Gut development: The regulation, evolution and function of *Bapx1*

Alison Julia Wright

Thesis Presented for the Degree of Doctor of Philosophy
University of Edinburgh
~ 2003 ~

Declaration

The experiments described in this thesis were the unaided work of the author except where acknowledgement is made by reference. No part of this work has been previously accepted for any other degree, nor is any part of it being submitted concurrently in candidature for another degree.

Alison J Wright

November 2003

Acknowledgements

The three years I have spent studying here in Edinburgh have been an incredible, and valuable, learning experience during which time I have met a number of wonderful people whose friendships I will always treasure. I am very grateful to Robert (Bob) Hill, for giving me the opportunity to work in his lab and for being such a supportive, caring and patient supervisor. I hope that some of your calm, relaxed approach to life has rubbed off on me! I am also delighted that you share my enthusiasm for the ski field and therefore understood the necessity of my annual disappearance to the slopes, not to mention various other sporting-related absences from the lab that have helped make my time here unforgettable!

There are numerous people at the HGU to whom I owe thanks for their help, advice and encouragement, without which this thesis would not have been possible. Amongst those are the fellow members of C3: Laura Lettice, who was able to answer virtually any scientific query I ever had; Rob Watson, who I am indebted to for accompanying me on those midnight dissection sections, and also for offering me a place on the ACE Race team which was one of the most exhilarating competitions I have ever taken part in (bring on the next one)!; Carlo de Angelis, whose lending-skills are greatly appreciated (how else would I have got here every day??); Simon Heaney, who is one of the nicest and most helpful people I have ever met; and Lorna Purdie for helping me with all of the transgenic work. I also greatly appreciate the help and advice of various members of the Hastie and Dorin lab: I am thankful for the encouragement provided by Heather MacPherson, and the kind offers to proofread thesis chapters from Lee Spraggon, who was also the only person capable of making me smile whilst I roasted in the flames of 'Thesis Hell'!! The presence of fellow students Duncan and Sapna during the late night thesis sessions was also somewhat comforting, although my introduction to the "Crap" was not so welcome!

I am very grateful to Paul Perry and Sandy Bruce in Photography, and various members of the PC support team, for their help in my times of computer-induced crises! I am also very grateful for the Bioinformatics advice provided by Colin Semple and Philippe (Gogo) Gautier, and to Duncan Davidson for the extended loan of his laptop on which to compose this thesis. I appreciate the histology help provided by Allyson Ross and the guidance from James Sharpe through the intricacies of OPT imaging (not to mention being indirectly responsible for the unprecedented growth of my CD collection … although how useful they will prove to be I'm not sure)! I am also grateful to all of the animal house staff for their help during my PhD.

There are many friends outside of the lab who I credit with making my time in Edinburgh such an enjoyable and unforgettable one. I couldn't have lived without the trampolining, tennis, dancing, swimming…etc. outlets! I am especially indebted to the constant encouragement from friends during the misery of 'Thesis Hell', who persisted with enquiries after me when it must have seemed as if I had disappeared from the face of the earth! Above all, I am indebted to my family for their unwavering and infallible source of love, support and understanding. No matter what the situation, you were always there when I needed you and it is often easy to forget the importance of that blessing. Having broken my chains to the library, I very much look forward to thanking you all in person in the near future. As testament to my enormous relief to be finally exiting 'Thesis Hell', I can't help but

adapt a quote from Claude Bernard, the father of experimental medicine (Bernard, 1865): "[L'achèvement de cette thèse] est un salon superbe tout resplendissant de lumière, dans lequel on ne peut parvenir qu'en passant par une longue et affreuse cuisine". ([The completion of this thesis] is a superb and resplendent salon that one can only reach by passing through a long and dreadful kitchen).

Abstract

The NK homeobox gene family comprises an extensive collection of transcription factors which have been shown to play pivotal roles in cell-fate specification and organogenesis. Murine *Bapx1* (*Nkx3.2*) was isolated as an evolutionarily conserved homologue of *Drosophila bagpipe* (*NK3*) which specifies the visceral mesoderm giving rise to gut musculature. A role for *Bapx1* in gut development has been conserved to mammals, where it would appear to be required both to specify the identity and to direct asymmetric growth of the spleno-pancreatic mesenchyme. Epithelial-mesenchymal signalling is an essential part of gut organogenesis, however the precise role of the mesenchyme is poorly understood and critical time points for its presence remain unknown. Initial investigations to probe the importance of the mesenchyme focussed on attempts to ablate this specific cell population by directing a toxic phenotype to *Bapx1*-expressing cells. A recombineering approach using *Bapx1*-containing PACs was employed to generate the targeting construct, however inability to modify the *Bapx1* locus necessitated adoption of an alternative approach.

The regulatory elements responsible for directing tissue-specific *Bapx1* expression have not yet been determined, therefore to pursue investigation of the role of the mesenchyme in gut development would require elucidation of a gut-specific *Bapx1* enhancer. Comparative sequence analysis between human, mouse and fish identified two candidate control elements, termed ProxB and DistB, which reside 11kb and 18kb respectively downstream of *Bapx1*. ProxB constitutes a short region of high sequence conservation between human and mouse, within which 180bp was also found to be conserved to zebrafish. DistB encompasses a more extensive 8kb genomic region highly conserved in mammals. Both the ProxB putative regulatory element and a 2kb subfragment of DistB named DistB1 were cloned into a *LacZ* reporter construct and assayed for enhancer activity in transgenic mice. Both elements exhibited tissue-specific enhancer capacities with ProxB-driven expression initially confined to the mandibular portion of the first branchial arch and later detected in Meckel's cartilage and the middle ear. Detailed analysis of the expression pattern and comparison to evolutionarily conserved expression domains provides good evidence that ProxB constitutes a region-specific *Bapx1* regulatory element. DistB1 appears to be capable of directing expression to several different tissues, notably including the digits where it closely resembles endogenous *Bapx1* expression. Further parallels between the DistB1 and *Bapx1* expression domains suggest that this element may also contribute to the endogenous regulation of *Bapx1*.

Phenotypic comparisons between *Bapx1* null mutant mice and their wildtype counterparts have contributed significantly to knowledge of the wildtype function of the gene. Within the developing gut, distinct molecular boundaries are apparent. *Bapx1* demarcates the domain of the posterior stomach and asplenic null mutants attest to its being essential for formation of the spleen. Various gut markers were employed to study the disruption of both gut development and defined molecular domains in the absence of *Bapx1*. A marked difference in *Sonic hedgehog* (*Shh*) expression was observed in mutant guts, revealing a fusion of the upper duodenum to the posterior stomach to contribute to the enlarged stomach phenotype. Furthermore, *Shh* expression was detected in an abnormal branch emanating from the duodenum and extending alongside the posterior stomach thus exposing a novel phenotypic manifestation of the mutation.

A novel anatomical structure, the splanchnic mesothelial plate (SMP) has been implicated in leftward mesenchymal growth during spleen morphogenesis. Cells within the SMP express *Bapx1*, without which leftward growth is disrupted and asplenia results. Prompted by these observations, a description of spleen development was undertaken to closely examine the precise origin and subsequent morphogenesis of this organ. Expression of the spleen marker *Hox11* was followed in dissected guts within a defined developmental window. The marked morphological transformation from a small domain of *Hox11*-expressing cells posterior to the stomach to an extensive layer flanking the dorsal mesogastrium is described herein. This work prompts questions as to whether a wave of cell-cell signalling or cell migration is responsible for the observed development of the spleen.

Table of Contents

List of Figures

List of Tables

List of Movies

List of Abbreviations

°C	degrees centigrade
A	absorbance
Amp	ampicillin
A-P	anterior-posterior
BAC	bacterial artificial chromosome
BMP	bone morphogenetic protein
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CNS	conserved noncoding sequence
DEPC	diethyl pyrocarbonate
dH_2O	distilled water
DMF	dimethyl formamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	dinucleotide triphosphate
D-V	dorsal-ventral
ds	double stranded
E	embryonic day
EDTA	ethyldiaminetetra-acetic acid di-sodium salt
EST	expressed sequence tag
EtOH	ethanol
FGF	fibroblast growth factor
FISH	fluorescence *in situ* hybridisation
g	gram
HCl	hydrogen chloride
HGU	Human Genetics Unit
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilo base pairs
L-R	left-right
l	litres
LacZ	*β-galactosidase* gene
LMP	low melting point
LPM	lateral plate mesoderm
M	molar
m	prefix *milli*
μ	prefix *micro*
Mb	mega base pairs
MeOH	methanol
mRNA	messenger RNA
myr	million years
n	prefix *nano*
ncRNA	noncoding ribonucleic acid
nt	nucleotides
OD	optical density

OPT	optical projection tomography
ORF	open reading frame
PAC	P1 artificial chromosome
PBS	phoshpate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
pmoles	pico moles
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse transcription-polymerase chain reaction
SDS	sodium dodecyl sulfate
SMP	splanchnic mesothelial plate
ss	single stranded
SSC	saline sodium citrate
TAE	tris, acetic acid, EDTA
TBE	tris, boric acid, EDTA
TE	tris, EDTA
TGF	transforming growth factor
UTR	untranslated region
UV	ultraviolet
Xgal	5-bromo-4-chloro-3-indolyl-β-**D**-galactoside
3D	three dimensional

Chapter 1 Introduction

1 Introduction

1.1 Gut development

The gut consists of a number of functionally distinct and highly specialised organs whose role within the digestive system is critical for the maintenance of all multicellular organisms. As the organism ingests food, it must be processed and essential nutrients absorbed into the body prior to expulsion of the waste products. The gut is an evolutionarily early adaptation and its developmental aspects are well conserved throughout the vertebrates. Figure 1.1 illustrates the formation of the digestive system in the early embryo. Despite its ultimate complexity, the vertebrate gut initially forms as a simple, relatively straight tube with few distinctive features along its anterior-posterior (A-P) axis. It is composed of two tissue types: a luminal lining of endoderm-derived epithelium surrounded by an overlying splanchnic mesoderm. Morphological observations attest to the similar formation of the embryonic gut along the A-P axis in all animals (Gilbert, 1997). The innermost layer formed during gastrulation is the endoderm, and the primitive tubular structure of the embryonic gut is created by two invaginations in the ventral walls of the gut endoderm at the rostral and caudal ends of the embryo. These invaginations, termed the anterior intestinal portal (AIP) and the caudal intestinal portal (CIP), migrate towards each other, the endoderm behind them forming two open-ended tubes. These tubes extend towards each other until they meet and fuse at the umbilicus (yolk stalk) of the embryo (Gruenwald, 1941).

In the vertebrate embryo, the mesoderm arises through inductive processes, and after gastrulation it is patterned along the dorsoventral and mediolateral axes to form axial, paraxial, intermediate and lateral plate mesoderm (LPM). The LPM further segregates into two components, somatic and splanchnic, and during formation of the early gut tube, splanchnic mesoderm is recruited to surround the invaginating endoderm. Cell populations within these mesodermal domains subsequently differentiate into individual mesodermal tissues including the mesenchyme of the internal organs, heart, spleen, cartilage, bones, smooth musculature and blood.

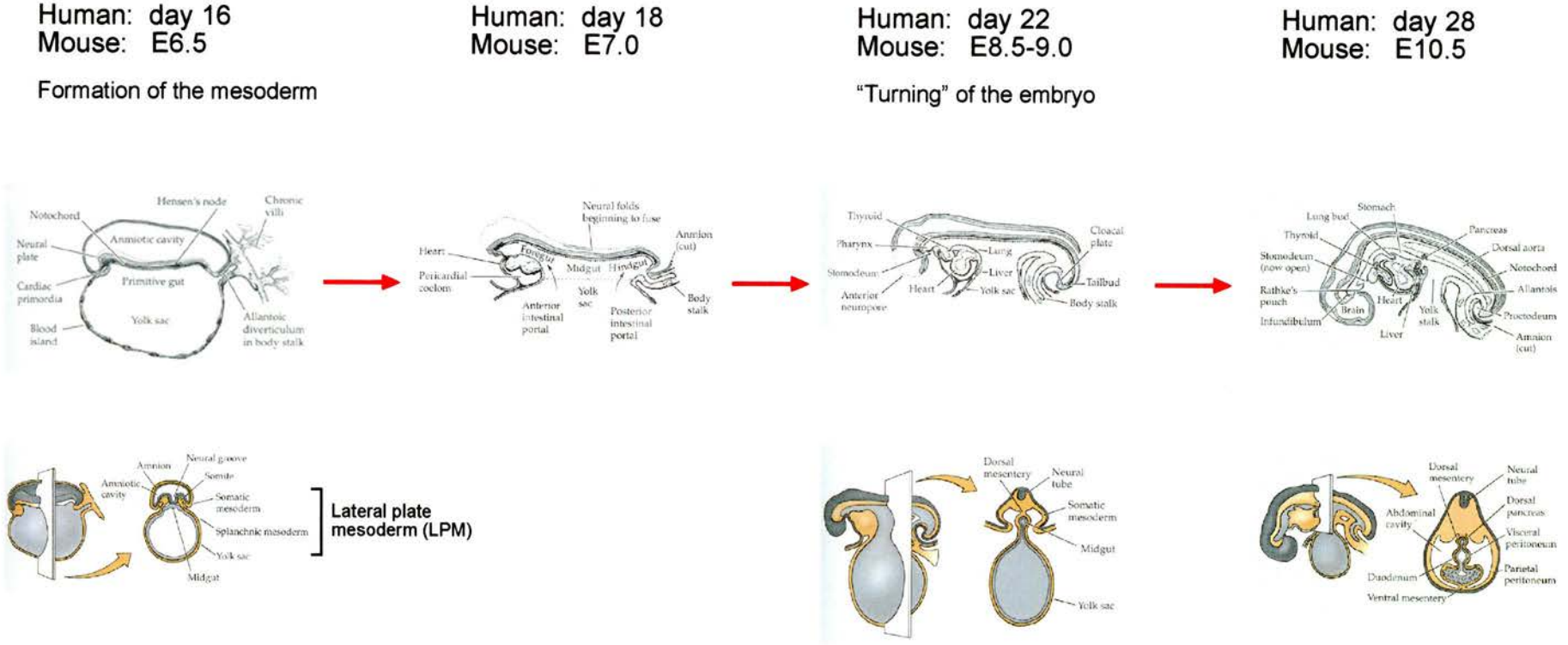


Figure 1.1: Formation of the digestive system in an early human embryo. Four successive developmental stages are illustrated, with the corresponding age of a mouse embryo indicated. Figure adapted from Gilbert, 1997.

As gut development proceeds, the uniform luminal epithelium undergoes regional specification along the A-P axis (Kedinger *et al.*, 1988). During subsequent morphogenesis and differentiation, the endoderm and mesoderm acquire their organ-specific characteristics, producing the organs of the gastrointestinal tract. Functionally and morphologically, the gut tube is subdivided into three regions: foregut, midgut and hindgut, for which distinct molecular boundaries exist. The foregut and hindgut are derived from the AIP and CIP respectively, whilst the midgut is composed of tissue originating from both the AIP and CIP (Carlson, 1999). The focus of this thesis is predominantly the foregut and, in particular, the spleno-pancreatic region in which I include the endodermally-derived pancreas, the anterior duodenum and the posterior part of the stomach together with the mesodermally-derived spleen. The anatomical proximity of the developing organs within the spleno-pancreatic region is illustrated by Figure 1.2.

1.1.1 The importance of the mesenchyme

Within the dynamic context of a developing organism, countless morphogenetic events take place as tissue organisation emerges. The cells of the mesenchyme, along with those that comprise epithelial sheets, are the two principal cell types participating in organogenesis.

1.1.1.1 Definition of the mesenchyme

The term *mesenchyme* is loosely defined but is generally used to describe embryonic cells that have not differentiated into a specialised cell type and are found as individuals or in groups, rather than in sheets. Since mesenchymal cells are usually found in groups, it is their social behaviour that is generally studied in the morphogenetic context. With regard to germ layers in the early embryo, the mesenchyme is not only comprised of derivatives of the mesoderm, but also includes

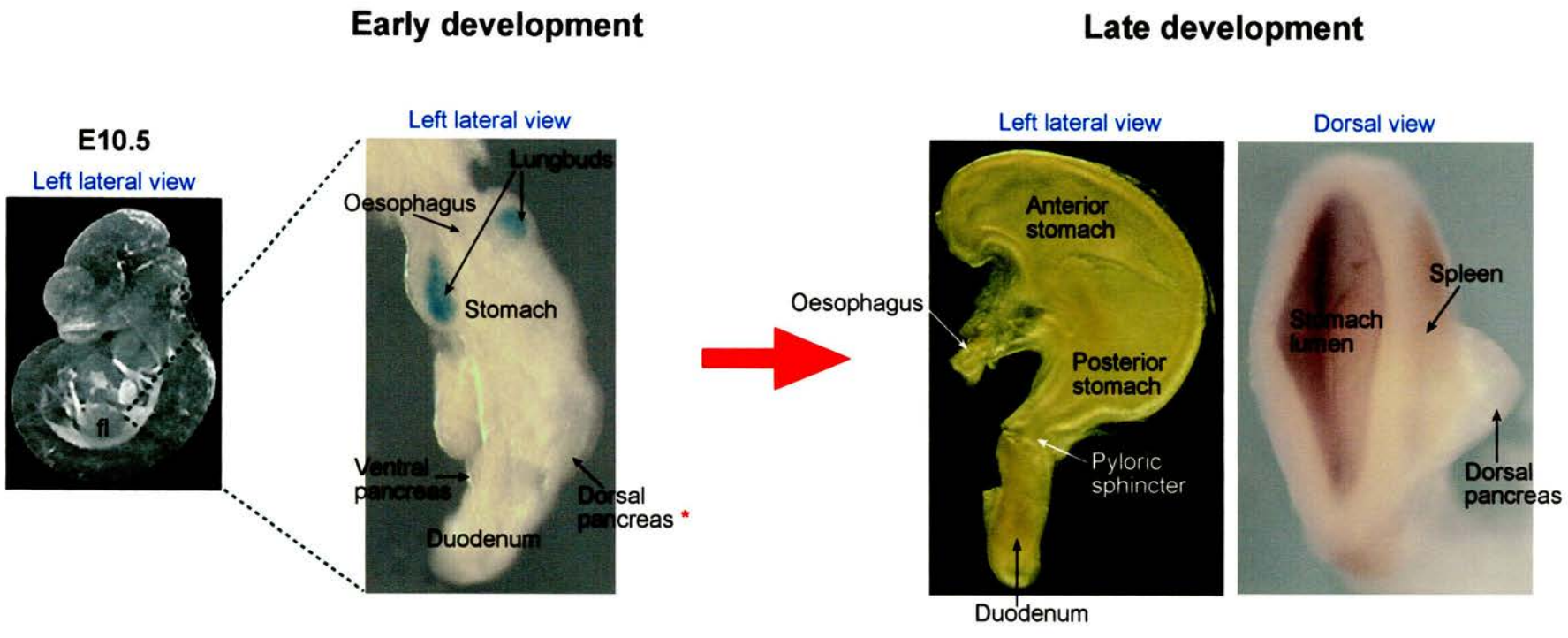


Figure 1.2: An illustration of foregut development in the embryonic mouse. Note the anatomical proximity of the developing organs at E10.5. The spleno-pancreatic region comprises the posterior stomach, duodenum, dorsal pancreas and spleen. The spleen is not evident as a separate structure at E10.5, at which point it is developing within the mesenchyme surrounding the dorsal pancreas (*). A *LacZ* reporter gene expressed in the lungbuds at E10.5 and detected by Xgal staining highlights the position of these organs. By later stages of embryogenesis, significant growth and development of the spleno-pancreatic region has occurred and both the dorsal pancreas and the spleen are positioned along the dorsal edge of the stomach. fl, forelimb.

contributions from ectodermal derivatives in the form of neural-crest cells (Bard, 1990).

1.1.1.2 Patterning properties of the mesenchyme

An important pattern-forming property of mesenchymal cells is their ability to move. Thus for organogenesis to proceed successfully, it is vital that this movement be constrained, directed and polarised by the environment and by neighbouring cells. Epithelial-mesenchymal signalling is a critical factor in many developmental processes and several studies have shown that reciprocal interactions are instrumental in the promotion of tissue growth and cytodifferentiation. Examples include tissues as diverse as the tooth (Miletich and Sharpe, 2003), lung (Have-Opbroek, 1991), kidney (Davies, 2002) and pancreas (Golosow and Grobstein, 1962). Furthermore, the formation of condensations, as seen for example during establishment of the axial skeleton and spleen (Thiel and Downey, 1921; Erlebacher *et al.*, 1995), is a common prelude to more complex organogenesis and is often the first stage in the formation of a wide variety of tissues. Hence the participation of mesenchymal cells in this morphogenetic condensing process attests to their significance during development.

1.1.1.3 Epithelial-mesenchymal interactions

Epithelial-mesenchymal signalling is a critical factor in many developmental processes and plays an important role in the specification of the embryonic gut. During the initial patterning of this organ, a primary signal emanating from the endoderm specifies the surrounding mesoderm to become gut mesoderm (Kedinger *et al.*, 1986, 1990) and the secreted molecule *Sonic hedgehog* (*Shh*) has been assigned an important role in this process (Roberts *et al.*, 1995; Apelqvist *et al.*, 1997). *Shh* is initially expressed uniformly throughout the gut epithelium with the exception of some associated glands and ducts, such as the trachea, liver, pancreas and epithelial-derived glands of the chick proventriculus (glandular stomach)

(Apelqvist *et al.*, 1997; Kim *et al.*, 1997; Narita *et al.*, 1998). By embryonic day (E)11.5, *Shh* has been significantly down-regulated in the posterior stomach (Bitgood and McMahon, 1995) and is no longer detectable (Kim *et al.*, 2000). Neither is it expressed in the pancreatic endoderm or the mesodermally-derived spleen.

The role of *Shh* in the morphogenesis and cytodifferentiation of the stomach has been extensively studied in the chick. The emerging picture is one of dynamic interactions subject to extensive spatial and temporal regulation, and for which the influence of the mesenchyme is a critical factor in multiple aspects of stomach development (Fukuda and Yasugi, 2002; Fukuda *et al.*, 2003). Studies in the mouse have demonstrated that exclusion of *Shh* expression from the pancreatic anlagen is essential for normal development of the pancreas and spleen (Apelqvist *et al.*, 1997; Hebrok *et al.*, 1998). Furthermore, the presence of temporally-specific pancreatic mesenchymal factors is necessary for the morphogenesis and cytodifferentiation of the pancreatic epithelium (Golosow and Grobstein, 1962; Wessels and Cohen, 1967). Thus endodermally-derived signals regionally specify the mesoderm, which in turn patterns the phenotype of the underlying endoderm. Conserved gene families that have been implicated in this reciprocal patterning process include the *Bone Morphogenetic Proteins* (*Bmps*) and *Hox* family of transcription factors (Roberts *et al.*, 1995, 1998; Smith and Tabin, 1999; Narita *et al.*, 2000), in addition to the role of *Shh* discussed above.

Epithelial-mesenchymal signalling pathways, together with their constituent signalling molecules, are beginning to be better understood. Various members of the *Bmp*, *Nkx* and *Wnt* families have been assigned important roles in signalling interactions mediating development of the foregut (Narita *et al.*, 2000; Smith *et al.*, 2000a; Theodosiou and Tabin, 2003). Further elucidation of factors involved in these interactions should be aided by the determination of expression profiles whereby molecular boundaries of gene expression are coincident with morphological boundaries of the gut.

1.1.2 Stomach development

Early A-P patterning of the primitive gut involves formation of the stomach, which demarcates the posterior boundary of the foregut and is first seen as a left-sided bulge in the middle of the anterior half of the gut tube (Gilbert, 1997). At E9.5 in mouse development the presumptive stomach can be visualised as a swelling of the endodermal tube at the level of the forelimbs. By E12.5, it is possible to detect morphological differences between the anterior (fore) and posterior (glandular) stomach; the former has a mucous lining whilst the latter is lined with a columnar epithelium resembling the vili in the gut (Hecksher-Sørensen, 2001). The passage of food from the stomach into the small intestine is regulated by the pyloric sphincter, which is an organ derived from the foregut (Smith and Tabin, 1999) and is distinguished by its thickened mesodermal layer. The pyloric sphincter thus marks the junction between the foregut and the midgut.

1.1.2.1 Evolution

Whilst the developmental aspects of the gut are well conserved throughout the vertebrates, the most obvious variation is evident in the stomach (Smith *et al.*, 2000b). The evolutionary origin of the vertebrate stomach is unknown. The primitve chordate *Amphioxus* does not possess a stomach and there exist four orders within the fish class also lacking this organ (Walker and Liem, 1994). The amphibian *Xenopus* is characterised by a metamorphosis that occurs between the larval stage and adulthood; the adult amphibian is considered to possess a true stomach, as do the remaining vertebrate lineages (Smith *et al.*, 2000b). Figure 1.3 compares the mouse gut to the equivalent structure in the chick. The avian stomach is composed of two different regions separated by a distinct constriction. The anterior portion of the stomach is called the proventriculus (glandular stomach), which develops compound glands and secretes pepsinogen, whereas the posterior portion of the stomach is termed the gizzard (muscular stomach), which develops a powerful muscular layer and carries out mechanical digestion (Fukuda and Yasugi,

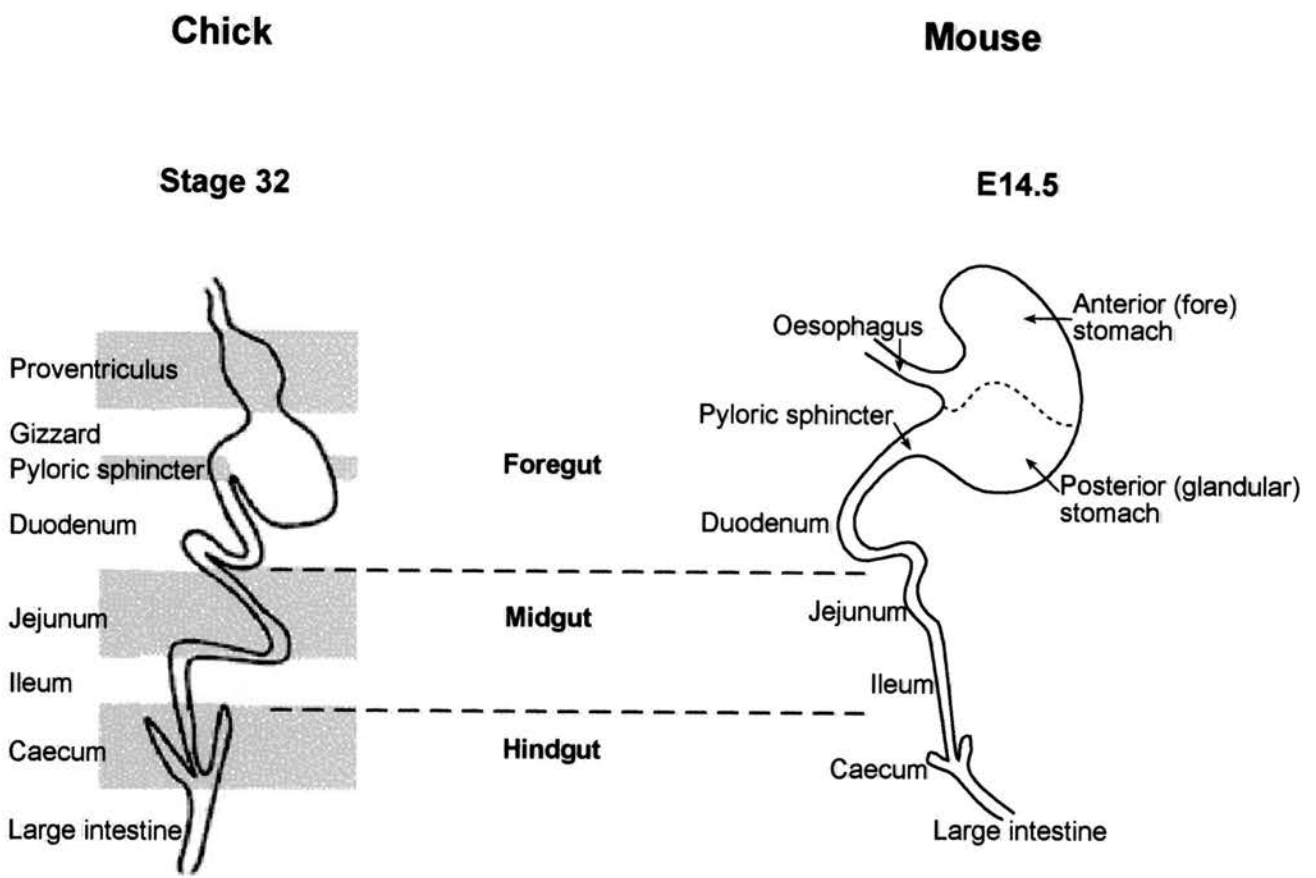


Figure 1.3: A comparison of the chick (*Gallus gallus*) and mouse (*Mus musculus*) gut. A comparable developmental stage is shown; stage 32 in chick (Hamburger and Hamilton, 1951) is the equivalent of E14.5 in mouse (Butler and Juurlink, 1987). The principal anatomical difference is the morphology of the stomach which in chick comprises two sacular structures: the proventriculus (glandular stomach) and an avian-specific adaptation, the muscular gizzard.

2002). The gizzard is an avian-specific adaptation that is thought to have evolved in the birds to replace mastication (Smith *et al.*, 2000b and references therein).

The phylogenetic relatedness of the amphibian, avian and mammalian stomachs has been investigated in terms of ontogenetic similarities and by the use of molecular criteria. Despite being separated by greater than 300 million years of evolution (Hedges and Kumar, 2002), the avian and mammalian stomachs were found to exhibit remarkably similar molecular and morphological phenotypes (Smith *et al.*, 2000b). Histological observations, together with gene expression patterns using molecular markers of the anterior and posterior stomach in addition to the small intestine, support the hypothesis that the anterior stomach of mice is homologous to the avian proventriculus, whilst the posterior stomach in mouse is homologous to the gizzard in chicken (Smith *et al.*, 2000b). In particular, the conserved expression patterns of genes, some of which have been assigned important roles in stomach development (discussed below), suggests that the patterning of the stomach is an ancient blueprint that was established before the divergence of the avian and mammalian lineages.

1.1.2.2 The importance of molecular boundaries

The amphibian and mammalian lineages have an estimated separation time exceeding 360 million years (Hedges and Kumar, 2002). The morphogenesis of *Xenopus* during its development from a tadpole to the adult frog requires that a morphologically and functionally distinct digestive system is first patterned for the larval phase. During the subsequent conversion from a herbivorous tadpole to a carnivorous frog, the conformation of the *Xenopus* stomach and intestine is completely remodelled (Hourdry *et al.*, 1996). Whilst the molecular markers *per se* appear to be conserved in *Xenopus*, their expression patterns in the embryonic gut were found to deviate from those conserved between birds and mammals (Smith *et al.*, 2000b). Notably, the lack of distinct gene expression boundaries in the *Xenopus* embryonic gut correlates with a lack of defined organ boundaries. In contrast, the chicken and mouse exhibit both distinct gene expression boundaries and distinct

organ boundaries (Smith *et al.*, 2000b). These findings suggest that the establishment of a unique molecular signature, as determined by the expression of a particular combination of genes, is an essential prerequisite during the specialisation of distinct regions of the gut.

The general conservation of molecular markers within the amphibian, avian and mammalian guts suggests that gut development may be based upon an ancient patterning event. However the variety of habitats occupied and great assortment of food types ingested by different species has resulted in a diverse array of shapes, sizes and types of stomach. A model that has been proposed to account for stomach patterning during vertebrate evolution involves slight modifications in temporal and spatial gene expression patterns leading to the appearance of specialised stomach compartments adapted to suit the needs of the given species (Smith *et al.*, 2000b). The refinement of gene expression patterns at some point during development is required to establish the boundaries within which specific organs will form.

1.1.2.3 Epithelial-mesenchymal interactions

The epithelial layer of the stomach derives from the endoderm, whereas the mesoderm contributes a layer of connective tissue and smooth muscle. At early stages, *Shh* is expressed throughout the endoderm, however its expression is later restricted to the anterior stomach (Kim *et al.*, 2000). Activin receptor-mediated signalling has been assigned a role in the restriction of *Shh* expression (Kim *et al.*, 2000). Both *ActRIIB-/-* and *ActRIIA+/-B-/-* mice, which contain mutations in Activin receptors, display a transformation of posterior stomach into anterior stomach, with stomach and pancreatic defects being more severe in the latter mutants (Kim *et al.*, 2000). Biochemical and genetic data suggest that ActRIIA and ActRIIB can interact with a number of ligands, notably including members of the BMP family (Yamashita *et al.*, 1995; Chang *et al.*, 1997; McPherron *et al.*, 1999; Oh *et al.*, 2002; Shi and Massague, 2003), amongst which BMP2 and BMP4 have important roles in stomach patterning (Roberts *et al.*, 1998; Smith and Tabin, 1999; Narita *et al.*, 2000).

Epithelial *Shh* has been shown to induce *Bmp* and *Hox* gene expression in the adjacent mesenchyme (Roberts *et al.*, 1995). However, whilst ectopic expression of *Shh* can induce *Bmp4* in the mesoderm of the midgut and hindgut, it is unable to do so in the gizzard (posterior stomach) mesoderm (Roberts *et al.*, 1998), suggesting that the mesoderm is prepatterned prior to the expression of *Shh*. Studies in chick have proposed that *Bapx1*, which is expressed in the posterior gizzard, plays a role in the regulation of *Bmp4* and *Wnt5a*. Ectopic expression of *Bapx1* in the anterior proventriculus results in transformation to a gizzard-like morphology along with inhibition of the normal proventricular expression of *Bmp4* and *Wnt5a* (Nielsen *et al.*, 2001). Crosstalk between the mesoderm and the endoderm evidently plays an important role in region-specific differentiation within the gut. For example, when *Hoxd-13*, normally restricted to the mesoderm of the posterior hindgut, is misexpressed in the primitive midgut mesoderm, it induces differentiation of the adjacent endoderm to acquire hindgut characteristics (Roberts *et al.*, 1998). The specification of the pyloric sphincter at the junction between the posterior stomach and anterior duodenum is a prime example of the precision with which epithelial-mesenchymal signalling must operate during development. Initial signals emanating from the endoderm are transduced by the mesoderm such that only a specific region of the latter expresses *Nkx2.5* and is thus competent to form the pyloric sphincter via signalling back to the endoderm (Smith *et al.*, 2000a).

1.1.3 Pancreas development

The pancreatic anlage is determined in mouse by E8.5, prior to the appearance of the pancreatic buds, which form as a dorsal and ventral evagination of the foregut endoderm directly posterior to the stomach and later fuse around E13.0-E14.0 (Wessells and Cohen, 1967). Figure 1.4 illustrates early stages in the development of the dorsal pancreas.

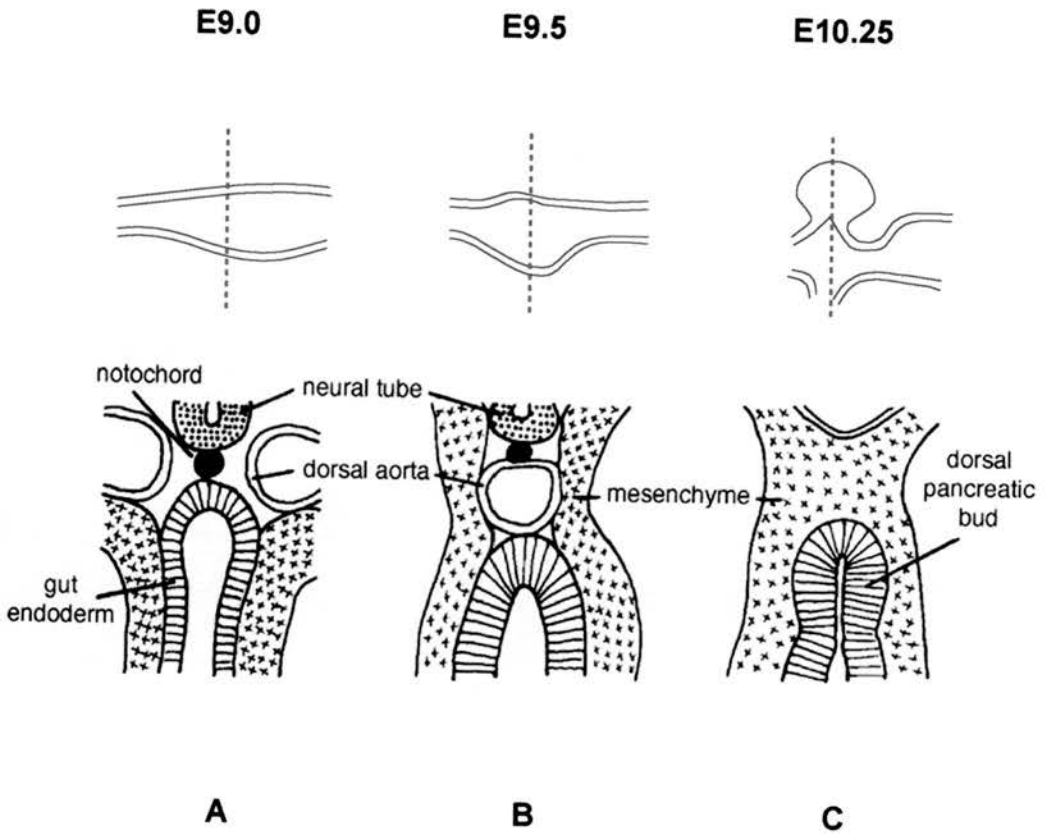


Figure 1.4: Early stages in the development of the dorsal pancreas in mouse. (**A**) At the 15 somite stage (E9.0), the notochord abuts the neural tube and the gut. (**B**) At the 20 somite stage (E9.5), the dorsal aorta lies between the gut and the notochord. (**C**) By the 28 somite stage (E10.25), mesenchyme surrounds the whole gut and the dorsal pancreatic bud has formed. Red dotted lines through the gut tube (outlined in brown) indicate the position and angle of the corresponding transverse sections which appear below. Taken from Slack, 1995.

1.1.3.1 Inductive interactions

Both the exocrine and the endocrine components of the pancreas are of endodermal origin (Percival and Slack, 1999). The pancreatic mesenchyme consists of loose cells of mesodermal origin and whilst they do not contribute to the exocrine or endocrine pancreas, the presence of the mesenchyme has been demonstrated to be essential for the morphogenesis and cytodifferentiation of the pancreatic epithelium (Golosow and Grobstein, 1962). Moreover, the action mediated by the mesenchyme could be exerted through a filter membrane, suggesting that the mesenchyme secreted a soluble growth factor needed by the epithelium. Investigations pertaining to the temporal specificity of the mesenchymal factor(s) established that the endoderm requires pancreatic mesenchyme to develop beyond a certain stage, after which heterologous mesenchyme is sufficient to promote pancreas development (Wessells and Cohen, 1967).

Appropriate differentiation of the pancreas relies upon signals from the notochord (Kim *et al.*, 1997) and from blood vessel endothelium (Lammert *et al.*, 2001), which both derive from the mesoderm. These signals are deduced to act upon endoderm that is already pre-patterned, since neither is capable of inducing pancreatic markers in endoderm that would not normally form pancreas. Recent progress has been made in the identification of the instructive signal through the use of chick-quail chimeras and *in vitro* tissue recombination. These investigations have assigned an important role for the LPM in sending instructive signals to the endoderm to induce pancreatic development (Kumar *et al.*, 2003). The instructive signals provided by the splanchnic mesoderm were capable of inducing pancreatic gene expression in non-pancreatic endoderm. These studies thus shed light on the way in which the initial domains within the endoderm are established *in vivo* and, notably, provide evidence that BMPs and Activin may be important signalling molecules in this process (Kumar *et al.*, 2003). Parallels can thus be drawn between the molecular pathways implicated in both stomach and pancreas specification (Kim *et al.*, 2000).

1.1.3.2 Expression of molecular markers

Studies to characterise the genes involved in regulating pancreas organogenesis have revealed distinct signalling pathways that interact to promote or restrict morphogenesis and cell differentiation. Failure to inhibit *Shh* in the endoderm at the pancreatic A-P level of normal mice appears to direct pancreatic mesenchyme cells to an intestinal fate (Apelqvist *et al.*, 1997). Activin, a TGFβ signalling molecule, has been implicated in the repression of *Shh* expression required to permit pancreas development (Hebrok *et al.*, 1998). Mutations in Activin receptors disrupt organogenesis of the stomach, pancreas and the spleen (Kim *et al.*, 2000). *Patched1* (*Ptc1*), encoding a repressor and negative regulator of *hedgehog* activity, and *Indian hedgehog* (*Ihh*) are expressed in pancreatic tissue and inactivation or mutation of these genes disrupts normal development and functioning of the pancreas respectively (Hebrok *et al.*, 2000).

Pancreatic and duodenal homeobox gene 1 (*Pdx1*) (originally isolated as *Insulin promoter factor 1* (*Ipf1*); Ohlsson *et al.*, 1993) and *Islet-1* (*Isl1*) both function in pancreas organogenesis and serve as useful markers during early development. *Pdx1* is required for morphogenesis of the pancreatic epithelium and also reportedly has a role in the progressive differentiation of the endocrine cells (Ahlgren *et al.*, 1996). It does not, however, affect the mesenchyme, demonstrating that morphogenesis of the latter is uncoupled from that of the pancreatic epithelium (Ahlgren *et al.*, 1996). In contrast, *Isl1* is required for the formation of the dorsal pancreatic mesenchyme and in its absence, there is an associated failure of exocrine cell differentiation in the dorsal pancreas (Ahlgren *et al.*, 1997). Additionally, functional *Isl1* in pancreatic endodermal cells is critical for the generation of all endocrine islet cells. *N-cadherin* constitutes another important gene involved in the pancreatic epithelial-mesenchymal signalling pathway; the dorsal pancreas fails to form in N-cadherin-deficient mice (Esni *et al.*, 2001). *N-cadherin* has been demonstrated to be required for the survival of the dorsal pancreatic mesenchyme and a proposed influence upon pancreatic morphogenesis involves signals from the mesenchyme promoting rearrangement of the cytoskeletal filament system within the endoderm, thus influencing both cell shape changes and growth (Esni *et al.*, 2001).

FGF10, a member of the fibroblast growth factor (FGF) family, constitutes a secreted mesenchymal growth factor acting early in pancreas development (Bhushan *et al.*, 2001). As embryogenesis proceeds, *Fgf10* is transiently expressed in pancreatic mesenchyme during the early stages of epithelial bud formation. Analysis of the pancreatic buds in *Fgf10-/-* mutant mice led to the proposal that *Fgf10* signalling regulates the proliferation of, and thus the size of, the *Pdx1*-expressing epithelial progenitor cell population (Bhushan *et al.*, 2001). Expression of *Fgf10* in the mesenchymal layer adjacent to the posterior stomach suggests that this signalling molecule could potentially play a similar role in the development of the stomach and/or the spleen. The spatial and temporal expression patterns of various signalling molecules within the mesenchyme are likely to be key factors in the ability of the mesenchyme to elicit different responses from the epithelium.

1.1.4 Spleen development

The spleen is an organ asymmetrically located along the left-right (L-R) axis that initially functions in haematopoiesis and later in immunity. There is a conspicuous paucity of data regarding the early formation of the spleen. This is despite the immediate proximity of the spleen and pancreas (see Figure 1.2) and indeed the development of the former from within the dorsal pancreatic mesenchyme.

1.1.4.1 Evolutionary perspective

Our closest invertebrate ancestor, *Amphioxus*, does not possess a spleen and the initial site of development of this organ in vertebrates was within the intestine (intra-enteric). The evolution of the vertebrates brought about a change in the location of the spleen from intra-enteric to extra-enteric sites. Thus phylogenetically, the spleen is classified as follows: 1) intra-enteric dispersed type in the hagfish, a primitive vertebrate, 2) intra-enteric diffuse type with aggregations within the spiral valve in the lamprey, a jawless vertebrate, 3) intra-enteric type with segregations within the wall of the stomach in the lungfish, a jawed vertebrate, and 4) extra-enteric type

attached to the dorsal mesentery (including the dorsal mesogastrium) and sharply segregated from the stomach or intestine in most other vertebrates (Saito, 1984 and references therein). The study by Saito pays particular attention to the relationship between spleen development and the gastro-enteric vasculature. This leads to a division of spleen ontogeny into four phases: phase I, appearance of a splenic primordium related to the gastroenteric arteries; phase II, development and proliferation of the splenic sinuses; phase III, formation of the splenic portal system; and phase IV, independence of the spleen from the gastroenteric canal. Phase IV is the final condition of the splenic phylogenetic history and, whilst observed in the adults of most higher vertebrates, not all of the lower vertebrates progress to this phase (Saito, 1984). Interestingly, the first three phases of spleen ontogeny have shortened considerably during vertebrate evolution.

1.1.4.2 Embryonic development

The initial appearance of the splenic rudiment as a mesenchymal condensation along the left side of the mesogastrium dorsal to the stomach has been noted in fish (Saito, 1984), amphibians (Manning and Horton, 1969), avians (Yassine *et al.*, 1989) and mammals (Thiel and Downey, 1921; Green, 1967). The mesenchymal cells of the splenic primordium accumulate and proliferate in the area of the dorsal surface of the dorsal pancreas, placing the origin of spleen development within or adjacent to the dorsal pancreatic mesenchyme. Subsequent growth and differentiation of the mesenchyme leads to formation of the splenic anlage, a structure recognisable as the spleen, which continues to exhibit considerable growth and extends intra- or extra-enterally depending upon the site of its initial appearance and the amount of available space at the site. In mice, the splenic anlage is normally apparent by E11.5 and by E13.5, the spleen has extended along a significant proportion of the dorsal mesogastrium flanking the stomach.

Prior to the appearance of the spleen, a transient, thick, columnar organisation of mesothelium lining one side of the foregut region has been observed (Green, 1967). This thickened mesothelial layer, named the anterior splanchnic mesothelial

plate (ASMP), persisted longest over the lateral portion of the pancreatic mesoderm in which the spleen develops. Interestingly, the ASMP was found to be reduced or absent in mice carrying a mutant allele of the *dominant hemimelia* (*Dh*) gene which, notably, are asplenic (Green, 1967). The correlation between absence of a spleen and lack of a thickened coelomic epithelium suggested a role for the ASMP in spleen development (Green, 1967). The thickened mesothelial layer particular to the level of the dorsal pancreatic bud will herein be referred to as the splanchnic mesothelial plate (SMP). Detailed studies of early spleen development, including factors involved in the specification of this organ have, until relatively recently, been limited by the availability of tissue-specific markers to identify splenic precursor cells.

1.1.4.3 The genetic basis of spleen development

In addition to the spontaneous mouse mutant *Dh*, a number of additional genes have been implicated in spleen organogenesis by virtue of targeted inactivation in mice. Null mutations in *Hox11*, *Bapx1*, *Capsulin*, *Wilms' Tumour 1* (*Wt1)* or *ActRIIB* result in asplenic mice (Roberts *et al.*, 1994; Oh and Li, 1997; Herzer *et al.*, 1999; Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000; Lu *et al.*, 2000) whilst *Nkx2.3* mutants exhibit either disrupted spleen development or, in some cases, asplenia (Pabst *et al.*, 1999). *Nkx2.5* constitutes an early marker for spleen precursor tissue (Patterson *et al.*, 2000) however targeted disruption of *Nkx2.5* in mice results in embryonic lethality between E9.0 and E10.0 (Lyons *et al.*, 1995) which precludes analysis of spleen development in these mutants. The arrest in embryogenesis is attributed to the failure of the heart tube to undergo complete looping, which is a critical process during cardiac development. The elucidation of evolutionary conserved homologues of *Hox11* and *Nkx2.5* (Logan *et al.*, 1998; Patterson and Kreig, 1999; Patterson *et al.*, 2000), which exhibit similar expression domains, suggests that, as in the case of stomach development, conserved pathways may exist that control development of the spleen.

Whilst a number of genes have thus been implicated in spleen organogenesis, determination of their precise function in this process remains the subject of

investigation. WT1 is expressed during mammalian embryonic development in many tissues and genetic analyses have clearly demonstrated the importance of WT1 for a large variety of developmental processes. Besides deficiencies in spleen development, homozygous *Wt1* null mice display gonadal and renal agenesis, and severe heart, lung, adrenal and mesothelial abnormalities (Kreidberg *et al.*, 1993; Herzer *et al.*, 1999; Moore *et al.*, 1999). The absence of the spleen, or existence of only a rudimentary spleen, in *Wt1-/-* knockout mice correlates with a dramatic increase in apoptotic activity which is also observed in many other organs such as the gonads, kidneys and retina (see review by Wagner *et al.*, 2003).

The dorsal pancreatic mesenchyme appears at approximately E9.0 in mouse (Slack, 1995) and expression of various splenic markers such as *Hox11* and *Nkx2.5* has been demonstrated at E10.5 in the splanchnic mesenchyme underlying the SMP (Hecksher-Sørensen *et al.*, 2003, submitted). Expression of *Hox11* is first observed in the dorsal mesogastrium at E10.5 whereas *Wt1* expression does not commence until E11.5 (Koehler *et al.*, 2000). Analysis of *Hox11* null mice demonstrated a significant reduction in *Wt1* mRNA levels and, together with evidence that *Hox11* could transactivate the *Wt1* promoter, led to the suggestion that *Hox11* is involved in the maintenance of normal levels of *Wt1* gene activity in the spleen (Koehler *et al.*, 2000). Another putative target gene of *Hox11* is *Aldh1*, which was discovered at elevated levels in the splenic anlage of *Hox11-/-* embryos (Greene *et al.*, 1998). Through regulation of *Aldh1* function in the spleen, *Hox11* may indirectly control retinoic acid availability and thus the asplenic phenotype of *Hox11* null mice may involve an apoptotic effect of excess retinoids (Greene *et al.*, 1998).

The defect in *Hox11*-deficient mice occurs after the initiation of spleen morphogenesis (Dear *et al.*, 1995). Thus, as in the *Capsulin* null mutants, splenic specification occurs however the primordium fails to develop beyond an initial group of precursor cells. Spleen development therefore appears to consist of genetically separable steps. In the case of *Capsulin* and *Wt1* null embryos, apoptosis is proposed to account for the resultant asplenia (Herzer *et al.*, 1999; Lu *et al.*, 2000) and a putative role proposed for *Capsulin* is to control morphogenetic expansion of the splenic anlage (Lu *et al.*, 2000). Insight into the effect exerted by *Hox11* on spleen development has been facilitated by the use of chimaeric mice consisting of wildtype

and *Hox11-/-* cells. These studies demonstrated that despite initial mixing of wildtype and mutant cells, a reorganisation subsequently occurred such that *Hox11-/-* cells were excluded from the developing spleen and formed a distinct cell population (Kanzler and Dear, 2001). Furthermore, as opposed to elimination by cell death, these mutant cells persisted in the dorsal mesogastrium as an unorganised rudiment. The inability of wildtype cells to rescue the defect provided further indication that *Hox11* acts cell autonomously to control differentiation of spleen precursor cells and that in its absence spleen development is arrested.

Evidence that the pancreas *per se* is not required for spleen development is provided by studies of mice lacking *Pdx1* or *Hlxb9*, in which the dorsal pancreas fails to form but the spleen develops normally (Jonsson *et al.*, 1994; Ahlgren *et al.*, 1996; Li *et al.*, 1999; Harrison *et al.*, 1999). Exclusion of *Shh* expression from the dorsal pancreatic endoderm (which prevents exposure of the splanchnic mesenchyme to *hedgehog* signalling) is an essential requirement for spleen induction as illustrated by failure of the spleen to form in *Pdx1-Shh* transgenic mice in which *Shh* expression is directed to the dorsal pancreatic endoderm (Apelqvist *et al.*, 1997).

1.1.4.4 Asymmetric growth of the spleno-pancreatic region

The processes involved in the generation of L-R asymmetries can be partitioned into three stages: orientation of the L-R axis relative to the A-P and D-V (dorsal-ventral) axes, asymmetric gene expression and asymmetric morphogenesis of specific organs. Prior to E8.0, the mouse embryo exhibits morphological symmetry and the presumptive spleno-pancreatic region is positioned between the future forelimb buds in alignment with the embryonic midline. The first morphological sign of L-R asymmetry is the rightward looping of the primitive heart tube at E8.5. The gut subsequently undergoes morphogenesis between E9.5-E12.0 whereby the linear gastrointestinal tube undergoes looping and rotation (at two main sites), resulting in the stomach, spleen and pancreas being asymmetrically positioned at the left side of the visceral cavity. (The current opinion regarding the establishment of L-R asymmetry is reviewed in depth by Hamada *et al.*, 2002).

Prior to gut rotation the endoderm is surrounded by mesoderm that is flanked by a mesothelial layer on each side of the anterior gut. As rotation proceeds, the mesothelium on the left side is maintained whilst that on the right becomes thinner, resulting in a thickened mesothelial layer termed the SMP residing on the left side of the gut mesoderm at E10.0 (Figure 1.5) (Hecksher-Sørensen *et al.*, 2003, submitted). The SMP is thickest at the level of the dorsal pancreatic bud and this transient embryonic structure is thought to have a role in asymmetric growth (Hecksher-Sørensen, 2001; Hecksher-Sørensen *et al.*, 2003, submitted). Moreover, recent work attests to a putative inductive role for the SMP in spleen development since the SMP is defective in both *Dh* and *Bapx1*-deficient mice in which the spleen is absent (Hecksher-Sørensen *et al.*, 2003, submitted).

A growing collection of genes which exhibit asymmetric expression along the L-R axis have now been identified (Patterson *et al.*, 2000). Amongst these, *Nkx2.5* and *Bapx1* initially demonstrate symmetric expression in the LPM but subsequently become restricted to the splanchnic mesenchyme on the left side (Patterson *et al.*, 2000; Hecksher-Sørensen *et al.*, 2003, submitted). Owing to its early expression in pre-splenic tissue, *Nkx2.5* has been employed to study the generation of spleen handedness. Investigations using *Nkx2.5*-expressing explants suggested that stage-specific tissue interactions are responsible for the downregulation of *Nkx2.5* on the right side of the embryo (Patterson *et al.*, 2000). Analysis of embryos with axis reversals provided evidence that the handedness of the heart and spleen are not linked whereas the direction of gut looping and location of the pancreas show a marked correlation with the side of *Nkx2.5* expression and hence location of the spleen (Patterson *et al.*, 2000). The observation that around E10.5, the pancreatic epithelium begins to invade the surrounding mesenchyme (Wessels and Cohen, 1967), attests to both the intimacy and complexity of spleen and pancreas development. Considering the spatial relationship between organs within the spleno-pancreatic region, it is possible that local growth signals (perhaps provided by the SMP) play an important part in the determination of L-R asymmetry of tissues.

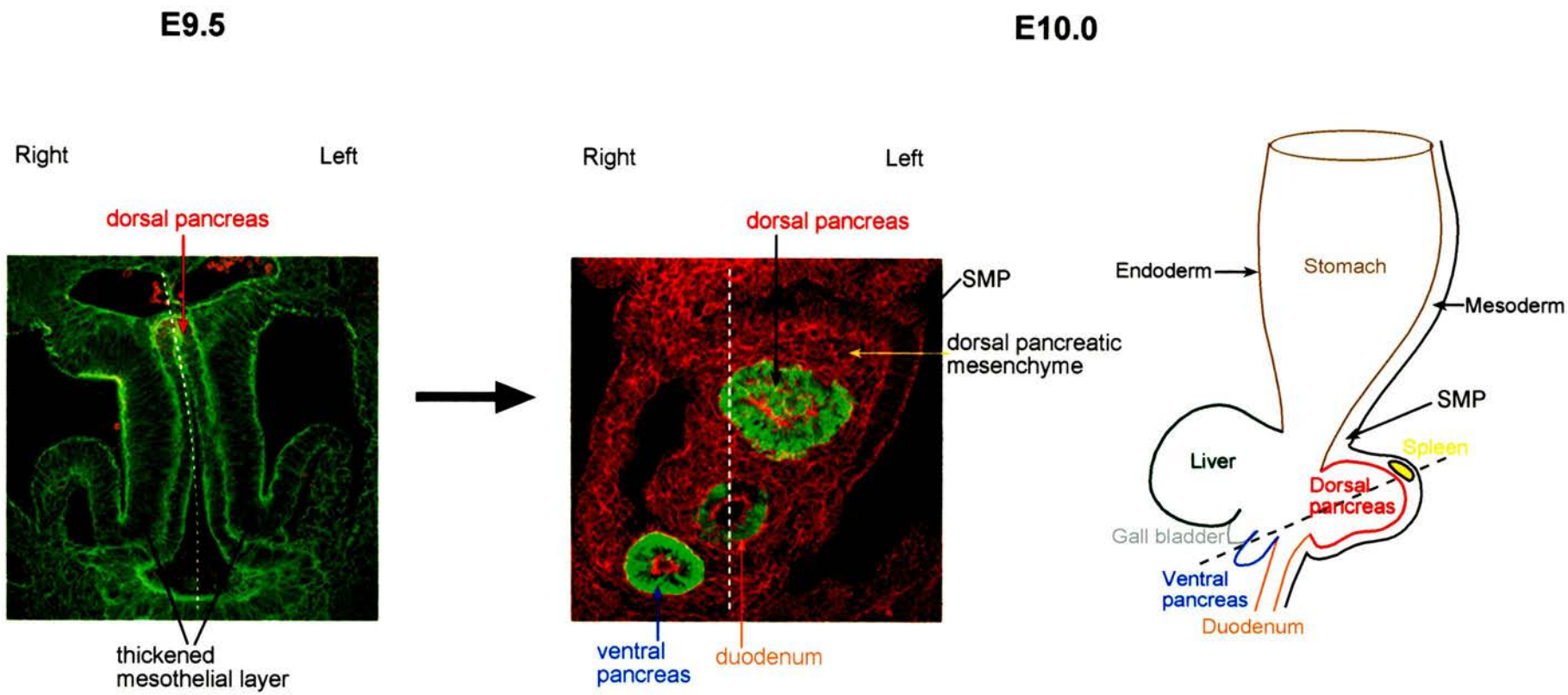


Figure 1.5: Formation of the splanchnic mesothelial plate (SMP) precedes the appearance of the spleen. Asymmetric growth of the spleno-pancreatic region between E9.5 and E10.0 is illustrated by transverse sections taken through the early gut. The thickened mesothelial layer on the left side of the foregut endoderm is maintained whilst that on the right side narrows. White dotted lines denote the embryonic midline. The black dotted line in the schematic indicates the corresponding section angle at E10.0. The mesodermally-derived spleen develops from within the dorsal pancreatic mesenchyme. Phalloidin (conjugated to a green secondary antibody at E9.5 and a red secondary antibody at E10.0) marks F-actin within cell walls whilst PDX1 (red at E9.5, green at E10.0) marks the pancreatic endoderm. Transverse sections were kindly provided by Dr. Jacob Hecksher-Sørensen.

1.2 *Bapx1*

1.2.1 Conservation, isolation and genomic structure

Murine *Bapx1* (*Bagpipe homeobox gene 1*), also known as *Nkx3.2*, is a member of the NK-2 homeobox gene family and was isolated as an evolutionarily conserved homologue of the *Drosophila bagpipe* homeobox gene (*NK3/bap*) which is involved in specification of the visceral mesoderm giving rise to gut musculature (Tribioli *et al.*, 1997). *NK* homeobox genes have been shown to play pivotal roles in cell-fate specification and differentiation of various organs. A human homologue, *BAPX1*, has also been isolated (Tribioli *et al.*, 1997) and assigned to chromosome 4p16.1 (Yoshiura and Murray, 1997), a locus to which several hereditary skeletal disorders have been mapped. *Bapx1* homologues have additionally been identified in fish (Miller *et al.*, 2003), amphibian (Newman *et al.*, 1997; Nicolas *et al.*, 1999) and avian (Schneider *et al.*, 1999) lineages. *Bapx1* resides at the proximal end of mouse chromosome 5, close to the *Msx1* gene (Tribioli *et al.*, 1997). It comprises two exons separated by a small (1.3kb) intron and encodes a predicted protein of 333 amino acids. The genomic organisation of *Bapx1* is depicted in Figure 1.6.

In addition to the highly conserved homeodomain, Bapx1 has a conserved region referred to as the TN domain, located near to the N-terminal portion of NK-2 class homeobox proteins (Harvey, 1996). There exists an additional conserved domain called the NK-2-specific domain (NK-2SD) located C-terminal to the homeobox (second exon) which has been identified in a number of NK superfamily genes (Harvey, 1996). Their precise function is unknown, although the TN domain of Nkx2.2 and Nkx6.1 shows sequence similarity to the core region of the engrailed homology-1 (eh1) domain present in the Engrailed (En) transcriptional repressor (Muhr *et al.*, 2001). It was subsequently demonstrated that Nkx2.2 and Nkx6.1 are able to interact with Gro/TLE corepressors via the TN domain (Muhr *et al.*, 2001). A final region of sequence identity comprising five amino acids has been observed C-terminal to the NK-2SD. It was named the Bap domain owing to its absence from any other NK-2 family member (Newman *et al.*, 1997).

1.2.2 Genomic regulation of *Bapx1*

The complete, annotated, genomic description of *Bapx1* at the nucleotide level, complete with extensive 5' and 3' flanking sequence, has not yet been established which hampers the process of identifying key regulatory elements controlling gene expression. Recently, transient transfection assays demonstrated the binding of two transcription factors, Pax1 and Pax9, to a motif located in the interval -880 to -844 relative to the +1 position which corresponds to the first nucleotide in the published cDNA sequence (GenBank Accession number U87957) (Rodrigo *et al.*, 2003) (Figure 1.6). No other Pax1/Pax9 binding sequences were revealed either in the genomic interval between –5285 to +109 or in the intron. Interestingly, transactivation by Pax1 and Pax9 was significantly enhanced in the absence of sequence between –5285 and –2762, suggesting the presence of *cis* elements with the ability to exert a repressive effect on these factors (Rodrigo *et al.*, 2003).

It is noteworthy that previous work in our lab involved the cloning of approximately 10kb of sequence from the *Bapx1* locus into a *LacZ* reporter construct (henceforth referred to as p10kb), including 7kb of sequence upstream of *Bapx1* exon1 (Figure 1.6). The p10kb vector was subsequently used to generate transgenic mice in order to assess the contribution of regulatory elements within this region and ultimately to attempt to define control sequences. Five transient transgenics were produced however the transgenic mice did not exhibit appropriate reporter gene expression as directed by p10kb, indicating the likely existence of *Bapx1* regulatory elements outside of the 7kb of sequence immediately upstream of the gene. In turn, this presents challenges regarding the elucidation of control sequences which may lie a significant distance from the gene itself, both upstream and/or downstream.

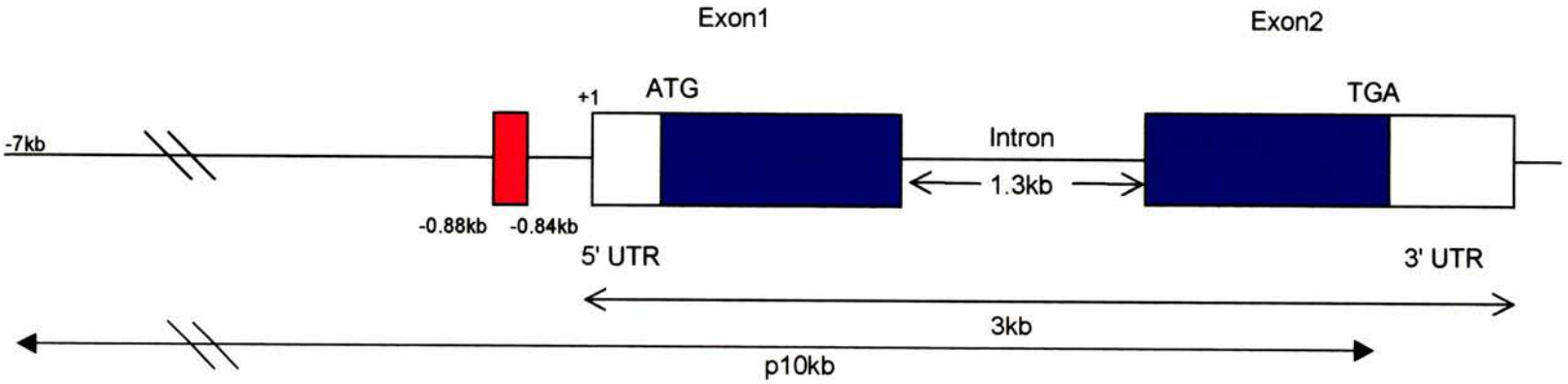


Figure 1.6: Genomic organisation of the *Bapx1* locus. Coding sequence is shaded blue and +1 indicates the 5' end of exon1. The proposed location of Pax1/Pax9 binding sites is shaded pink. Genomic sequence included in the p10kb vector is indicated.

1.2.3 *Bapx1* expression pattern during mouse development

1.2.3.1 The axial skeleton

Bapx1 is first detectable in embryos at E8.0-E8.5, just prior to axis rotation, in LPM adjacent to the endodermal lining of the prospective gut, and faintly in the second or third most recently condensed somites within the region corresponding to the presclerotome, the precursor of vertebrae (Tribioli *et al.*, 1997). Somites become polarised into a ventral mesenchymal compartment, the sclerotome, and a dorsal epithelial dermomyotome which gives rise to the skeletal muscles of the body and the dorsal dermis (Ben-Yair *et al.*, 2003). Along with *Pax1*, *Pax9* and *MFH-1*, *Bapx1* is regarded as one of the earliest markers for the sclerotome lineage that forms the axial skeleton (Deutsch *et al.*, 1988; Miura *et al.*, 1993; Neubuser *et al.*, 1995; Tribioli *et al.*, 1997; Akazawa *et al.*, 2000). *Bapx1* is expressed from the early stages of chondrogenesis, constituting a developmental marker for prechondrogenic cells that will subsequently undergo mesenchymal condensation, cartilage production and ultimately endochondral bone formation. Accordingly, by E14.5, transcripts are detected in the prevertebral column and at E16.5, expression within the axial skeleton is strongest in the most caudal prevertebrae, having decreased substantially in more rostral prevertebrae (Tribioli *et al.*, 1997).

1.2.3.2 The appendicular skeleton: limbs

At E10.5, *Bapx1* is detected in limb mesenchyme and by E12.5 it is expressed in all cartilaginous condensations of the developing limbs. By E14.5, transcripts are visible in the developing digits of the fore- and hindlimbs and at E16.5 *Bapx1* expression is strongest in the most distal cartilaginous condensations (namely the digits), decreasing in a distoproximal gradient (Tribioli *et al.*, 1997).

1.2.3.3 Craniofacial expression

At E10.5, *Bapx1* is detected in a discrete domain within the mandibular portion of the first branchial arch (Tribioli *et al.*, 1997). Regarding the latter, expression is seen by E12.5 in a domain associated with Meckel's cartilage which has recently been shown to include the primordia of the gonium and tympanic ring (Tucker *et al.*, 2003, submitted), both middle ear-associated bones. At E14.5, *Bapx1* expression was also observed in the developing malleus and incus which comprise, along with the stapes, the middle ear ossicles. Although these skeletal elements, with the exception of the stapes, are unique to mammals, the homologous bones in non-mammalian jawed vertebrates comprise the proximal jaw bones that form the jaw articulation (jaw joint). Notably, craniofacial expression of *Bapx1* has been reported in *Xenopus*, encompassing the jaw joint region (Newman *et al.*, 1997) and similarly, expression in cells within and surrounding the jaw joint has been detected in fish (Miller *et al.*, 2003). In the chick also, strong expression of *Bapx1* is observed initially in specific parts of Meckel's cartilage and the jaw joint region, later becoming downregulated in the cartilages but remaining high in the joint region (Tucker *et al.*, 2003, submitted). Thus it appears that the jaw/middle ear comprise a conserved site of *Bapx1* expression.

1.2.3.4 Gut expression

As mentioned above, *Bapx1* is detectable by E8.5 in LPM adjacent to the endoderm of the prospective gut. In *Drosophila* too, expression is apparent in gut mesoderm (Azpiazu and Frasch, 1993) therefore the latter would appear to represent an evolutionarily conserved site of expression between *Drosophila* and the higher vertebrates. As the splanchnic mesodermal tissues differentiate, *Bapx1* is detected in the mesoderm surrounding the posterior stomach. *Xenopus Xbap* has similarly been reported in the posterior foregut (Newman *et al.*, 1997) and in the chick, *Bapx1* is found in the more posterior of the two sacular structures that comprise the avian stomach, namely the gizzard (see Figure 1.3) (Nielsen *et al.*, 2001). As in the mouse, chicken *Bapx1* exhibits distinct molecular boundaries within the gut, being absent

from the pyloric sphincter at the posterior-most end of the stomach (Nielsen *et al.*, 2001).

At E9.5, *Bapx1* expression signals are detected in the mesothelium and underlying mesoderm on each side of the pancreatic endoderm (Hecksher-Sørensen *et al.*, 2003, submitted). However, between E9.5 and E10.0, the spleno-pancreatic region undergoes a dramatic leftward expansion coinciding with the formation of the SMP and the expression of *Bapx1*, initially symmetric, becomes restricted to the left side (Hecksher-Sørensen *et al.*, 2003, submitted). *Bapx1* is detectable throughout the SMP and in the underlying spleno-pancreatic mesenchyme at E10.25. Slightly later in development, at E10.5, expression is downregulated in the dorsal SMP but maintained in the spleno-pancreatic mesenchyme (Hecksher-Sørensen *et al.*, 2003, submitted) which later corresponds to *Bapx1* expression in the spleen.

There is now growing evidence that *Bapx1* has a role in the determination of L-R asymmetry. Although transcripts reportedly accumulate on opposite sides of mouse and chick embryos, in each case *Bapx1* expression is suggested to be subject to control by the L-R signalling pathways (Schneider *et al.*, 1999). A role for *Bapx1* in the development of the characteristic gut L-R asymmetry has been suggested in chick (Nielsen *et al.*, 2001) and, more specifically, through its role in the SMP, has been proposed in mouse (Hecksher-Sørensen *et al.*, 2003, submitted).

1.2.4 The acquisition of additional developmental roles during evolution

Following the initial isolation of *Bapx1* as a homologue of *Drosophila bap*, together with a conserved site of expression in the gut mesoderm, a possible evolutionary link between the two genes was suggested (Tribioli *et al.*, 1997). In *Drosophila*, the role of *bap* is in the specification of the visceral mesoderm that gives rise to gut musculature (Azpiazu and Frasch, 1993). This specific function within the gut would not appear to be conserved since no effect on the formation of gut musculature was observed in *Bapx1*-deficient mice (Lettice *et al.*, 1999). Nevertheless, despite differing consequences on cognate tissue types, the ancient role for *Bapx1* would

appear to be its function in certain aspects of gut development in both vertebrates and invertebrates.

The evolutionary appearance of the neural crest coincides with the appearance of the vertebrate phylum, being found solely in this group of animals (Gans and Northcutt, 1983). It has thus been suggested that, during evolution of the neural crest, the *bap* sequence was co-opted for use in specifying distinct elements of the jaw, as seen in zebrafish and *Xenopus* (Newman *et al.*, 1997; Miller *et al.*, 2003). With the evolution of mammals, the homologous jaw joint bones became the middle ear bones, and the role of *Bapx1* in the formation of these skeletal elements appears to have been conserved (Tucker *et al.*, 2003, submitted). A dramatic difference in the *Bapx1* expression pattern between mouse and *Xenopus* is its absence from the developing axial and appendicular skeleton of amphibians (Newman *et al.*, 1997). Thus, a further, novel developmental role for *Bapx1* appears to be its function in skeletal patterning in mammals. It is possible, however, that functions of the mammalian *Nkx3.2* gene are divided between more than one *Nkx3*-related gene in amphibians. In this regard, *Nkx3.3* (*zampogna*) (Newman and Krieg, 1999) has, to date, only been found in amphibians. Interestingly, the existence of 'job sharing' *Pax6a* and *Pax6b* genes is currently under investigation in zebrafish (personal communication; Prof. Veronica van Heyningen). The nature and relative importance of molecular changes to the *Bapx1* gene that have both permitted the acquisition of novel developmental roles and, potentially, facilitated evolution, constitute fascinating but unresolved questions. Figure 1.7 provides a summary of evolutionarily conserved sites of *Bapx1* expression.

1.2.4.1 Insights from phylogenetic studies

Nkx3-related genes have been isolated from a number of vertebrates and phylogenetic comparisons have demonstrated that they cluster into three subgroups, *Nkx3.1*, *Nkx3.2* and *Nkx3.3* (Lettice *et al.*, 2001). The divergence of the subgroups outside of the homeodomain suggests that they may have arisen via an ancient gene duplication event from a common ancestral gene (Lettice *et al.*, 2001). Interestingly,

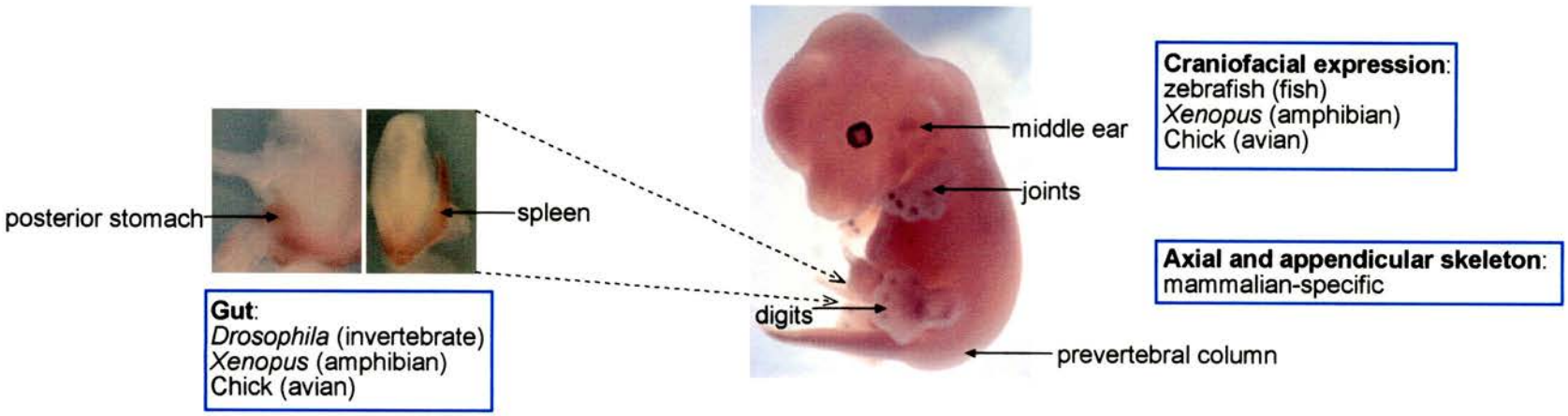


Figure 1.7: Regions of *Bapx1* expression in the embryonic mouse at E13.5. Conserved sites of *Bapx1* expression in evolutionarily diverged organisms are indicated. The righthand image was kindly provided by Dr. Robert Watson.

as mentioned above, homologues of *Nkx3.3* have only been identified in *Xenopus* (Newman and Kreig, 1999) and the urodele *Pleurodeles walt1* (Nicolas *et al.*, 1999) to date, thus it may be that *Nkx3.3* arose as a unique amphibian *Nkx3* member. Alternatively, as additional sequence from the fish and birds becomes available, it may transpire that *Nkx3.3* was originally present, but subsequently lost, in the evolutionary lineage leading to mammals. *Xenopus Nkx3.3* (known as *zampogna* (*zax*); Newman and Kreig, 1999) is expressed in the midgut musculature which is reminiscent of the expression domain of *Drosophila bap*. The genetic diversification of *bagpipe*-related genes could have arisen during gene duplication events that are predicted to have occurred twice in vertebrate evolution (Holland *et al.*, 1994). The *Bapx1* (*Nkx3.2*) gene is proposed to have acquired vertebrate specific functions at some point after the split between the jawless fish and the gnathostomes (vertebrate animals possessing a true jaw) (Lettice *et al.*, 2001). Subsequent functional divergence would permit the evolution of developmental innovations including those that are mammalian-specific.

1.2.5 The *Bapx1* mutant phenotype reveals specific developmental roles

The heterozygous *Bapx1*+/- genotype does not appear to adversely affect mice whose principal phenotypic manifestation upon birth appears to be a kinked tail (Lettice *et al.*, 1999). Haploinsufficiency of *Bapx1* may mildly affect some axial skeletal elements (Tribioli and Lufkin, 1999) and the malleus and gonium in the middle ear (Tucker *et al.*, 2003, submitted), but otherwise heterozygotes are viable, fertile, healthy and born in appropriate Mendelian ratios. This is in stark contrast to the homozygous *Bapx1*-/- mutants which die soon after birth and exhibit skeletal dysplasia, asplenia and gastroduodenal malformation (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000). The perinatal lethality of the *Bapx1* deficiency thus precludes *in vivo* analysis at postnatal stages.

1.2.5.1 Skeletal defects

The most profound defects are observed in the axial skeleton which derives from the sclerotome cells (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000). The distribution of these cells is affected by the loss of *Bapx1*; a markedly reduced number appear ventromedially around the notochord (Akazawa *et al.*, 2000). Subsequently, the intervertebral discs and centra of the vertebral bodies fail to form (Lettice *et al.*, 1999). The sclerotome cells of the mutants appear to migrate and condense normally, there being no alterations in cell number or density surrounding the notochord at E12.5 (Tribioli and Lufkin, 1999). *Bapx1* function was thus concluded to be dispensable for this process (Akazawa *et al.*, 2000). However subsequent proliferation is markedly reduced and differentiation fails to occur as demonstrated by the downregulation of a number of transcription factors, signalling molecules and extracellular matrix molecules which play a critical part in normal development of the foetal skeleton (Tribioli and Lufkin, 1999).

Pax1 and *Pax9* act synergistically in the proper formation of the vertebral column (Peters *et al.*, 1999) and their expression in sclerotome cells is not affected in the *Bapx1* mutants (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000), whereas *Bapx1* expression is lost in *Pax1/Pax9* double mutants (Rodrigo *et al.*, 2003). Together with data suggesting that Pax1/Pax9 bind to genomic sequence upstream of *Bapx1*, as mentioned above, *Bapx1* is thought to constitute a direct target of *Pax1* and *Pax9* (Rodrigo *et al.*, 2003). The vertebrae of *Bapx1*-/- foetuses appear to be more closely spaced along the entire vertebral column, this being particularly pronounced in the cervical region (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000). It is interesting to note that a role in skeletal development for the mouse paralogue of *Bapx1* (*Nkx3.2*), *Nkx3.1*, was revealed by the creation of homozygous *Nkx3.1/Nkx3.2* (*Bapx1*) double mutants, for which the resultant sclerotomal defects were enhanced as compared to homozygous *Bapx1* mice (Herbrand *et al.*, 2002). Abnormalities in *Bapx1*-/- mice were also found in the basioccipital and basisphenoid bones which are located at the base of the skull and derive from mesenchyme associated with the notochord. The conformation of the notochord is also reportedly abnormal, yet *Bapx1* is not detectably expressed in this

structure therefore the disruption was suggested to result from the lack of surrounding *Bapx1*-expressing sclerotomal cells (Lettice *et al.*, 1999).

Intriguingly there is no apparent phenotype in the limb, perhaps suggestive of functional redundancy in this region.

1.2.5.2 A role for *Bapx1* in craniofacial development

Gene targeting via a 'knock-in' of the *LacZ* gene to disrupt the *Bapx1* locus and create a null allele was employed to analyse the developmental functions of *Bapx1* (Akazawa *et al.*, 2000). In addition to sites previously reported in mice (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999), it is noteworthy that the reporter gene was expressed in the ribs and the mandible (lower jaw). Expression in the latter is reminiscent of the association of *Bapx1* with Meckel's cartilage and the jaw joint in fish, amphibians and avians (discussed above). The role of *Bapx1* in specification of the jaw joint has been demonstrated in zebrafish by the injection of morpholino antisense oligos (MOs) directed against *Bapx1* (Miller *et al.*, 2003). Injected animals displayed highly penetrant, dose-dependent loss of the jaw joint in addition to defects of the cartilaginous retroarticular process (RAP), which develops at the posterior end of Meckel's cartilage, and the retroarticular bone (RAB), which forms on the RAP of Meckel's cartilage (Miller *et al.*, 2003). Two markers of developing tetrapod limb joints, *Gdf5* (which encodes a TGFβ-related signalling molecule) and *chordin* (a BMP antagonist) were identified as downstream targets of *Bapx1* in the developing jaw joint owing to a marked reduction or absence respectively, specific to the joint region, in *Bapx1*-MO-injected fish (Miller *et al.*, 2003). Furthermore, expansion of the *Bapx1* expression domain into the ventral pharangeal cartilage of the lower jaw in *hand2* mutants led to the suggestion that repression of *Bapx1* in this region by *hand2* aids in the positioning of the jaw joint.

Regarding the mammalian equivalent of the jaw joint, detailed analysis of *Bapx1*-/- embryos has revealed a role for *Bapx1* in defining the structural components of the mouse middle ear (Tucker *et al.*, 2003, submitted). The role of *Bapx1* appears to have changed, however, during the evolution of the malleus and

incus (which constitute the homologues of the quadrate and articular in fish) since a comparable articulation defect is not seen in *Bapx1*-/- mice (Tucker *et al.*, 2003, submitted). Rather, the width of the malleus is decreased and defects are observed in both the gonium and the tympanic ring. Furthermore, contrary to the situation in fish, *Gdf5* expression in the mammalian joint region is not affected by the loss of *Bapx1* (Tucker *et al.*, 2003, submitted). Thus it has been postulated that the alteration in the role of *Bapx1* can be attributed, at least in part, to a modified regulatory potential with respect to downstream effectors involved in joint formation.

1.2.5.3 Gut defects

Bapx1-deficient mice are asplenic, with the absence of spleen precursor cells being observed from the earliest stages of initiation of the splenic anlage from within the splanchnic mesenchyme (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999; Akazawa *et al.*, 2000). The expression of *Hox11* and *Nkx2.5*, which are early markers of splenic precursor cells, is not induced in mutant mice (Lettice *et al.*, 1999; Hecksher-Sørensen, 2001). This demonstrates that the *Bapx1* gene is fundamental to early spleen development. Examination of homozygous *Bapx1* mutants demonstrated that loss of gene expression in the spleno-pancreatic mesenchyme was coincident with a defective SMP. At E10.5, this structure had failed to thicken in the mutant mice and left-sided growth was impeded, with the dorsal pancreas remaining along the embryonic midline (Hecksher-Sørensen *et al.*, 2003, submitted). Immunohistochemistry analysis confirmed the presence of the expected endocrine and exocrine cell types, therefore pancreatic cell differentiation still appears to proceed in the absence of the normal leftward growth (manuscript in preparation). *Fgf10* expression, however, was no longer detectable in the ventral SMP in the absence of *Bapx1* (Hecksher-Sørensen *et al.*, 2003, submitted) although the significance of this observation has not yet been determined. Nevertheless, on the basis of the aforementioned observations, it has been proposed that the asplenic phenotype may result from obliteration of a spleen-inducing signal from the SMP (Hecksher-Sørensen *et al.*, 2003, submitted).

Gastroduodenal malformation has also been noted in *Bapx1*-deficient mice, including loss of the pyloric sphincter at the posterior end of the stomach (Akazawa *et al.*, 2000). *Nkx2.5*, in addition to expression in the spleen, is also expressed in the pyloric sphincter and, in chick, it has been demonstrated to play a role in the patterning of the sphincter (Smith *et al.*, 2000a). The loss of *Nkx2.5* expression in *Bapx1*-/- mice thus presumably accounts, at least partially, for the concomitant absence of the pyloric sphincter.

Preliminary studies have uncovered novel manifestations of the *Bapx1* mutation in the spleno-pancreatic region involving morphogenesis of the pancreas and A-P boundary demarcation in the gastrointestinal tract. Regarding the former, the data indicates that *Bapx1* is required to confer functional identity to the spleno-pancreatic mesenchyme since in its absence, the dorsal pancreatic epithelium can be re-specified to acquire intestinal fates (manuscript in preparation). The dorsal pancreatic (splanchnic) mesenchyme appears to be either rescued by the latter and thus similarly exhibit characteristics of intestinal tissue, or die. Assays using various endodermal markers of the pancreas suggest that the posterior boundary of the stomach in *Bapx1* mutant mice has extended in a posterior direction along the A-P axis (manuscript in preparation). This gives the appearance of the pancreas budding out from the posterior stomach at E10.5, as opposed to emerging from a clearly distinct region of the anterior duodenum.

Explanation of the *Bapx1*-mediated abnormalities in gut development detailed above awaits further experimental analyses to probe *Bapx1* gene expression, regulation and function. The common existence of various anatomical and molecular boundaries within the spleno-pancreatic region of evolutionarily diverged vertebrates attests to the importance of gut development as a model system for understanding development. Furthermore, conservation of certain sites of *Bapx1* expression together with the emergence of novel expression domains during the course of evolution highlights the potential use of this gene as a tool to explore the relationship between molecular genetic changes and their associated morphogenetic effects.

1.3 Thesis Aims

The aims of this thesis broadly encompass:

To establish a means by which we can explore the role of the mesenchyme in gut development through the identification, and subsequent employment, of regulatory elements responsible for the tissue-specific *Bapx1* expression domains (Chapters 3 and 4).

To investigate the role of *Bapx1* in the specification of molecular boundaries within the foregut (Chapter 5).

To produce a 3 dimensional illustration of early morphogenesis of the spleen (Chapter 6).

Chapter 2 Materials and Methods

2 Materials and Methods

2.1 Manipulation of nucleic acids

2.1.1 General reagents used in molecular biology procedures

All chemicals were analytical grade and were supplied by Sigma, Roche, Fisher, Invitrogen and BDH. Restriction enzymes were supplied by Roche and New England BioLabs (NEB). Nucleic acid manipulations were performed in 1.5ml centrifuge tubes unless stated otherwise. General solutions were prepared by HGU technical staff, autoclaved and stored at room temperature.

Tris.HCl

Tris base (tris[hydroxymethyl]aminomethane) was dissolved in sterile water. HCl was used to adjust the pH to the required value.

EDTA

EDTA (ethyldiaminetetra-acetic acid di-sodium salt) was dissolved in sterile water. The solution was adjusted to pH8.0 by adding solid NaOH.

TE buffer

10mM Tris.HCl (pH7.5); 1mM EDTA.

TBE buffer, 20X stock

Tris base	216g
Boric acid	110g
0.5M EDTA	80ml

Distilled water was added to a final volume of 1 litre. Stock was diluted to 1X with distilled water.

TAE buffer, 50X stock

Tris base	242g
Glacial acetic acid	57.1ml
0.5M EDTA	100ml

Distilled water was added to a final volume of 1 litre. Stock was diluted to 1X with distilled water.

2.1.1.1 DEPC-treatment of solutions

Sterile water (dH_2O), phosphate buffered saline (PBS) and saline sodium citrate (SSC) were DEPC-treated by the addition of 1μl/ml DEPC (Sigma) followed by gentle mixing using a magnetic stirrer for 1-2 hours. The solutions were then allowed to sit at room temperature in a fumehood overnight prior to re-autoclaving.

2.1.2 Restriction enzyme digestion

Between 0.1μg and 5μg DNA was digested in a volume of 20μl, or 100μl for the larger amounts of DNA substrate. Enzyme was added at a concentration of 5-10u/μg DNA, depending on the duration of the digests. The manufacturer's guidelines were followed to determine the correct 10X buffer and incubation temperature for each enzyme. The final volume was achieved by addition of dH_2O.

2.1.2.1 Removal of buffer salts

In cases where multiple restriction digestion steps were performed and the enzymes required different buffers, digests were done separately. Buffer salts were removed between digests by drop dialysis and new buffers were added. A nitro-cellulose filter (Millipore) was placed in a petri dish containing dH_2O. The digestion reaction was carefully placed on the filter (a maximum of 50μl per filter) and salts were left to diffuse through the filter for at least 30 minutes at room temperature. Following

dialysis, the liquid was removed to a fresh 1.5ml centrifuge tube and the appropriate enzyme and buffer were added for the subsequent reaction.

2.1.3 Electrophoresis

2.1.3.1 Standard gel electrophoresis

DNA was size fractionated by agarose gel electrophoresis. To prepare the gel, the required amount of standard (High Pure, BioGene) or low melting point (LMP) (Invitrogen) agarose was dissolved in either 1XTBE or 1XTAE (weight/volume) by heating. Molten agarose was cooled and ethidium bromide added to a final concentration of 50μg/100ml agarose. DNA size markers were run alongside experimental samples to estimate the size and amount of DNA in the samples. The size marker used routinely was 1Kb DNA ladder (GibcoBRL). Samples were run in 1X loading buffer.

2.1.3.2 Agarose gel loading buffer

Loading buffer was prepared as a 10X stock as detailed below and stored at room temperature.

	Final concentration
Ficoll (Sigma)	15%
Orange G (Sigma)	1%

Made to required volume with TE, pH8.0.

2.1.4 Determining the concentration of DNA and RNA samples

DNA concentrations were determined in one of two ways: by agarose gel electrophoresis or by measuring DNA absorbance in a spectrophotometer at 260nm

(A_{260}). To determine the concentration by electrophoresis, several different volumes of the DNA were run alongside standard amounts of DNA (usually DNA size markers). An estimate of the concentration was made by visual comparison of the samples with the known amounts of DNA under UV illumination.

To determine the concentration by spectrophotometry, the DNA and RNA samples were diluted with DEPC-dH_2O. The spectrophotometer was calibrated using a water only blank sample. The samples were placed in clean cuvettes and the absorbance at 260nm (A_{260}) was measured. The concentration of DNA and RNA at a single A_{260} unit is given as:

1 A_{260} unit of dsDNA = 50μg/ml H_2O

1 A_{260} unit of ssRNA = 40μg/ml H_2O

The concentration of the original sample in μg/ml is calculated as follows:

Concentration (μg/ml) = A_{260} value x dilution factor x 50μg/ml for DNA; 40μg/ml for RNA.

2.1.5 Southern blot hybridisation

2.1.5.1 Production of radiolabelled probes

High specific-activity probes were generated by random priming using a Random Primed Labelling kit (Roche) and Redivue™ [α-^{32}P]-dCTP, 3,000 Ci/mmol (Amersham pharmacia biotech) according to the manufacturer's instructions. Template DNA derived from DistB1 PCR products. Unincorporated nucleotides were removed from labelled probes by column chromatography using Sephadex G-50 columns (Amersham pharmacia biotech) equilibrated with TE buffer. Probes were denatured by boiling for 5 minutes before addition to the hybridisation mix. Oligonucleotides (DistB: homology block 3: 5' AAGAAGGTGGTCCTAGTTGC 3'; homology block 8: 5' GGCCACATTCCAGTTTGTGC 3') were end-labelled using an End

Labelling kit (Roche) and Redivue™ [γ-^{32}P]-dATP, 5,000 Ci/mmol (Amersham pharmacia biotech) according to the manufacturer's instructions.

2.1.5.2 Immobilisation of DNA onto filters

Following restriction enzyme digestion and electrophoresis of samples, DNA was transferred onto Hybond-N+ membrane (Amersham pharmacia biotech) using standard techniques. Filters were then pre-hybridised in [0.5M Na_2PO_4 pH7.2, 7% SDS]. Denatured probe was added to the filters and hybridisation allowed to proceed at 65°C (48°C for end-labelled oligonucleotides) overnight prior to washing under increasingly stringent conditions in 2XSSC, 1XSSC, 0.1XSSC/0.1% SDS at 65°C (48°C for end-labelled oligonucleotides). Filters were exposed to X-OMAT AR film (Kodak), using intensifying screens, at -70°C for between 2 hours and 2 days.

2.1.6 DNA purification

This was accomplished by gel extraction and/or phenol-chloroform purification. For gel extraction, linearised DNA fragments or PCR products were electrophoresed using low melting point agarose prior to recovery of the DNA using a gel extraction kit (QIAGEN), according to the manufacturer's instructions. All optional steps were included.

Prior to phenol-chloroform purification, the DNA solution was re-suspended in sterile distilled water to a final volume of 250µl. The sample was then mixed with 250µl phenol:chloroform:isoamyl alcohol (25:24:1), vortexed for 30 seconds and subsequently centrifuged for 5 minutes at 13,000rpm. The upper aqueous layer was transferred to a clean tube and 250µl chloroform added, before vortexing for 30 seconds and centrifuging as before. The supernatant was again removed to a clean tube and DNA precipitated by the addition of 25µl (one tenth of the volume) 3M sodium acetate and 500µl (two volumes) ethanol (EtOH), followed by storage at

–20°C for 2 hours. The solution was centrifuged for 15 minutes at 4°C, after which the supernatant was discarded and the sample washed in 70% EtOH. The DNA was then air dried and finally re-suspended in dH_2O.

2.1.7 Ligations

2.1.7.1 Alkaline phosphatase treatment of vector DNA

Prior to ligation, vector DNA was treated with calf intestine alkaline phosphatase (CIP) (Roche) to dephosphorylate the 5' termini of vector DNA, preventing self-ligation. The linearised, purified vector DNA was made up to a volume of 90μl with sterile dH_2O followed by the addition of 10μl 10X buffer and 1μl CIP. Following 2x15 minute incubations, the first at 37°C and the second at 56°C, a further 1μl CIP was added and the 2x15 minute incubations repeated. The reaction was terminated by the addition of 2μl EDTA and the CIP enzyme heat inactivated by incubation for 20 minutes at 65°C. The dephosphorylated vector DNA was purified by phenol-chloroform extraction.

2.1.7.2 Ligation

In order to maximise the ligation efficiency, ligations were set up with 1:3 and 1:5 vector:insert molar ratios. Ligation controls lacking either insert or ligase were also included. The amount of vector in each case was approximately 20ng and the total reaction volume was 10μl. Ligations were carried out at 16°C overnight using T4 DNA ligase and ligation buffer (Roche). The ligase was inactivated the following morning by heating the ligation mixture to 65°C for 20 minutes. Prior to electroporation, salt from the ligation reaction was removed by filtration of the solution as described in section 2.1.2.1.

2.2 Microbiology

Aseptic technique was observed for all steps involving the growth of bacterial cells (setting up cultures, pouring agar plates, selecting single colonies and storing bacterial stocks). Liquid cultures were grown with vigorous shaking (220rpm) and dry cultures were grown on inverted agar plates at 37°C overnight.

2.2.1 Growth media for bacterial cultures

HGU technical staff prepared Luria-Bertani broth (L-Broth) and Luria-Bertani agar (L-agar) as detailed below:

L-Broth

Amount per litre:

Tryptone	10.0g
Yeast extract	5.0g
NaCl	10.0g
Glucose	1.0g

To prepare solid media, 15.0g agar was added to the solution prior to autoclaving.

2.2.1.1 Antibiotic selection

	Stock concentration	Working concentration
Ampicillin (Amp)	100mg/ml in dH_2O	100μg/ml

Aliquots were stored at -20°C.

2.2.1.2 Xgal/IPTG indicator plates

Before pouring the plates, Xgal was added to the molten L-agar at a final concentration of 40μg/ml whilst IPTG was added at a final concentration of 0.2mM.

2.2.2 Production of electrocompetent cells

A single colony of DH5α *E.coli* cells from an agar plate was used to inoculate approximately 10ml L-Broth for overnight growth at 37°C. The overnight culture was used to inoculate 2 X 400ml fresh L-Broth the following morning and cells were grown to an OD_{600} of 0.5-1.0. Flasks were chilled on ice for 15-30 minutes and cells were centrifuged at 4°C, 4,000g (6,000rpm) for 15 minutes. Pellets were re-suspended in 800ml ice cold sterile dH_2O. The centrifugation step was repeated and cells were re-suspended in 400ml chilled, sterile dH_2O. Following a further centrifugation step, cells were re-suspended in 20ml 10% cold, sterile glycerol. Cells were centrifuged once more and re-suspended in 2-3ml 10% glycerol. Aliquots of cells were frozen on dry ice and stored at -70°C.

2.2.3 Transformations

Competent cells were transformed with 10-100ng plasmid DNA by electroporation. Electrocompetent cells were thawed on ice and 50μl were added to each DNA sample in an ice cold centrifuge tube. The mixture was transferred to ice cold 0.2cm electroporation cuvettes (EquiBio) and allowed to sit for 1 minute on ice. Cells were electroporated using a BioRad Gene Pulser set at 25μF, 2.5kV, and 200Ω. Immediately after electroporation 1ml L-Broth was added to the cells, which were then allowed to recover for 1 hour at 37°C, shaking, prior to plating on selective media and overnight incubation at 37°C.

2.2.3.1 Identification of transformants

Modified DNA was identified by conducting a DNA preparation (section 2.2.4) prior to restriction enzyme digestion (section 2.1.2) and/or PCR analysis (section 2.3).

2.2.4 Isolation of DNA

2.2.4.1 Plasmid DNA

Plasmid DNA was prepared using a commercially available kit (QIAGEN). For the extraction of small amounts of plasmid DNA (minipreps), a single colony was selected to inoculate 2ml L-Broth containing a suitable antibiotic for overnight growth at 37°C. The following morning, plasmid DNA was extracted using the kit according to the manufacturer's instructions. Plasmid DNA was eluted in either 30μl or 50μl elution buffer. The DNA concentration was determined either by agarose gel electrophoresis or by spectrophotometry (section 2.1.4).

For the extraction of larger amounts of plasmid DNA (maxipreps), a 2ml starter culture was established as described above and incubated at 37°C for 8 hours. The starter culture was then re-suspended in 400ml L-Broth, with antibiotic selection, in a large conical flask to allow aeration of the bacteria and the larger volume culture was incubated at 37°C overnight. Plasmid DNA was extracted the following morning using the QIAGEN plasmid maxi kit according to the manufacturer's instructions. DNA was eluted in 400μl autoclaved TE, pH8.0 and the concentration determined as for miniprep DNA.

2.2.4.2 BAC DNA

BAC DNA was harvested essentially as detailed for large scale preparation of plasmid DNA (maxiprep; QIAGEN). Samples were subjected to alkaline lysis, the lysate cleared and DNA subsequently precipitated by addition of 0.6 volumes isopropanol. Samples were then centrifuged at 4,000rpm for 15 minutes at 4°C and

DNA pellets washed in 70% EtOH, air-dried and gently re-suspended in 400μl sterile dH_2O prior to phenol-chloroform extraction (section 2.1.6).

2.3 Polymerase chain reaction (PCR)

2.3.1 Reagents

dNTPs:

dNTPs were purchased from ABgene as stocks of 100mM. Working stocks of 25mM were made by mixing 25μl of each of the dNTPs (dATP, dCTP, dGTP, dTTP) to a final volume of 100μl and stocks were stored at –20°C. dNTPs were used in PCRs at a final concentration of 0.2mM.

Oligonucleotide primers:

PCR primers were designed by selecting a sequence, usually between 20 and 25 nucleotides in length, for which the [G+C]:[A+T] ratio was approximately equal. Care was taken to avoid the likelihood of primers self-annealing or annealing with each other. Longer primers were required when incorporating the T7 RNA polymerase promoter into the 5' end of the reverse primer, and also for certain targeting constructs. Primers were purchased from MWG Biotech as lyophilised desalted compounds. Stocks were made up at 100μM using sterile dH_2O and primers were used in PCRs at a final concentration of 0.5-1μM.

Additional reagents:

Ampli*Taq* DNA polymerase (5u/μl) (Roche) was used at 0.2μl per 20μl reaction. PCR buffer (Roche) was a 10X stock and therefore diluted 1:10 for reactions. Mg^{2+} (Roche) was used at a final concentration of 2.5mM unless stated otherwise. The reaction was made up to the required volume with sterile dH_2O. PCR amplification at the *Bapx1* locus required the addition of DMSO (Sigma) at a final concentration of

10% in order to enhance the efficiency of the PCR reaction by decreasing the stability of GC bonds and thus lowering the melting temperature. Routine PCRs were performed in 0.5ml centrifuge tubes in a MJ Research DNA Engine Tetrad.

2.3.2 PCR amplification programmes

General:

94°C for 4 minutes

94°C for 20 seconds

Annealing temperature for 20 seconds

72°C for 1 minute per kb of target sequence

Return to step 2 34 times

72°C for 5 minutes

Touchdown:

95°C for 5 minutes

93°C for 30 seconds

Annealing temperature for 50 seconds

70°C for 1 minute per kb of target sequence

Return to step 2 4 times

Repeat steps 2 to 4 5 times, each cycle reducing the annealing temperature (step 3) by 1°C

Repeat steps 2 to 4 29 times, maintaining the new annealing temperature

70°C for 10 minutes

2.3.3 Molecular cloning of PCR products

2.3.3.1 Sub-cloning via T-Easy vector (Promega)

Taq DNA polymerase normally adds a single nontemplated nucleotide (nearly always A) to the 3' end of all duplex DNA strands. Accordingly, PCR products were recovered from an agarose gel (section 2.1.6) and the DNA was subsequently re-tailed with dATP and ligated to T-Easy vector according to the manufacturer's instructions. 3μl of the ligation was transformed into either DH5α or BL21 cells. The BL21 cells were required to overcome the methylation sensitivity of *Asp*718 in the case of ProxB sequence (see Chapters 3 and 4). ProxB contains a recognition site for the *E.coli* Dcm methylase (a site-specific methylase encoded by the *dcm* gene) which overlaps the endonuclease restriction site, resulting in protection of DNA from cleavage. BL21 cells (a different *E.coli* strain) are *dcm*-, not harbouring the Dcm methylase, thus removing the block to restriction digestion.

2.3.3.2 Screening for transformants

Cells containing the T-Easy vector were selected for by their acquisition of Amp resistance. Insertion of sequence into the T-Easy vector results in disruption of the *LacZ* gene, thus transformants were identified as white, as opposed to blue, colonies when plated on IPTG/Xgal indicator plates.

2.4 Sequencing

Plasmid DNA used for sequencing was prepared using a QIAGEN miniprep kit according to the manufacturer's instructions. DNA was sequenced using the dye-labelled terminator dRhodamine for cycle sequencing. Reagents were thawed on ice and dyes were protected from light as much as possible. Reactions (20μl final volume) were set up in 0.2ml centrifuge tubes as follows:

200-500ng plasmid DNA in dH_2O	11μl
dRhodamine reaction mix	8μl
primer (3.2 pmoles)*	1μl

*The primers used were as follows:

m13 forward: GTAAAACGACGGCCAGT

m13 reverse: GGAAACAGCTATGACCATG

Cycle sequencing was performed using an MJ Research DNA Engine Tetrad. The sequencing program was [96°C, 30 seconds; 50°C, 15 seconds; 60°C, 4 minutes] for 24 cycles per reaction. Reactions were EtOH precipitated following transfer to a fresh 1.5ml centrifuge tube containing 50μl 100% EtOH and 2μl sodium acetate, pH5.2. Reactions were left on ice for 15 minutes to precipitate, before centrifuging at 13,000rpm for 30 minutes at 4°C. Pellets were washed with 200μl 70% EtOH. The supernatant was removed following a second centrifugation step and pellets were allowed to dry at room temperature for approximately 20 minutes. Samples were submitted to the sequencing service to be run on an ABI machine.

2.5 Animal husbandry

Animals used during transgenic procedures were maintained in a barrier facility and all other animals were maintained in a semi-barrier unit. All experiments were carried out under Home Office licence. Wildtype animals (CD1, CBA and C57BL/6) were obtained from Charles River Laboratories. Embryos for all experiments were generated from timed matings, with the morning of vaginal plug detection being considered as embryonic day 0.5 (E0.5).

2.5.1 Genotyping of breeding mice

Breeding animals were genotyped by PCR amplification of allelic sequences from genomic DNA extracted from ear clips. The DNA was extracted from each ear clip by incubating the tissue in 50μl of 25mM NaOH, 0.2mM EDTA at 95°C for 20 minutes, followed by vortexing briefly to break up the tissue. The reactions were neutralised by the addition of 50μl of 40mM Tris.HCl. For long-term storage, protein and cell debris were removed by centrifugation and decanting of the supernatant into fresh tubes. 1μl of extracted DNA solution was used in a 20μl *LacZ* genotyping PCR assay for which the primer sequences are given in Table 2.1. Extracted genomic DNA was stored at 4°C.

2.5.2 Harvesting of postimplantation embryos

Postimplantation embryos were harvested for *in situ* hybridisation studies and for *LacZ* staining. The mothers were sacrificed by cervical dislocation. The abdominal cavity was opened and uteri were removed to petri dishes containing DEPC-PBS. Embryos were removed from the uteri and freed from extraembryonic membranes using scissors and forceps. Extraembryonic membranes were retained separately to genotype embryos by PCR. Embryos were rinsed in fresh PBS prior to subsequent processing.

2.5.2.1 Genotyping of embryos

Embryos processed for *in situ* hybridisation and *LacZ* staining were genotyped using DNA derived from extraembryonic membranes. The latter were digested overnight at 55°C in 0.5ml yolksac lysis buffer containing proteinase K.

Yolksac lysis buffer:

	Final concentration
Tris.HCl pH8.0	100mM
EDTA	50mM
NaCl	100mM
SDS	1%

Proteinase K (stock 10mg/ml) (Sigma) was added to a final concentration of 0.2mg/ml.

Following overnight digestion, tubes were centrifuged at 13,000rpm for 10 minutes to pellet debris. The supernatant was transferred to a clean tube, to which 1ml chilled 70% acetone/5% DMF (in 25% dH_2O) was added. The solution was mixed manually by inversion of tubes and DNA pelleted by centrifugation for 10 minutes at 13,000rpm. The supernatant was then carefully decanted and DNA washed in 70% EtOH on a rotator for at least 30 minutes. Tubes were then centrifuged for 2 minutes at 13,000rpm, the supernatant discarded, and following air drying of samples, DNA was re-suspended in 100-500μl TE, pH8.0 depending upon the size of the pellet. Incubation at 55°C for 15 minutes aided re-suspension of the DNA.

PCR genotyping assays:

PCR amplification was performed using 1μl of the yolksac DNA. *Bapx1* and *LacZ* primer sequences and annealing temperatures are detailed in Table 2.1. The genotype was determined by agarose gel electrophoresis of the PCR product.

Visual assessment of the *Shh* mutation:

Wildtype and homozygous mutant *Shh* littermates aged E11.5, E12.5 and E13.5 were visually distinguished owing to gross morphological differences between the embryos.

2.5.2.2 Staging embryos by somite number

Animals used for the investigation of spleen development (Chapter 6) derived from C57/BL6xCBA crosses. Embryos were dissected in DEPC-PBS and staged by somite number according to the Theiler staging system (Theiler, 1972, reprinted 1989 with minor changes) as detailed in The Atlas of Mouse Development (Kaufman, 1992).

2.5.3 Microscopy

Dissections of post-implantation embryos were effected with the aid of a Leica Stereo MZFLIII stereo fluorescence microscope (Leica Microsystems, Milton Keynes, UK) fitted with a 100W Hg source for fluorescence imaging, and fibre optic cold light source and illuminated base for incident/transmission illumination. Images were captured with a Photometrics CoolSnap colour CCD camera (Roper Scientific, Tucson, Arizona) controlled by scripts written for IPLab Spectrum (Scanalytics Inc., Fairfax, VA).

2.5.4 Production of transgenic mice

2.5.4.1 Preparation of linearised recombinant DNA for microinjection into fertilised oocytes

Plasmid DNA harvested from T-Easy vectors identified as containing the desired sequence (section 2.3.3.2) was cleaved with *Asp*718 to release the desired PCR product prior to cloning into the *Asp*718-cleaved *LacZ* reporter construct (a derivative of plasmid pBGZ40; Yee and Rigby, 1993). Plasmid DNA was digested (5μg DNA in a 100μl reaction) with enzymes to release the transgene. Each digest was electrophoresed on a 0.8% LMP agarose gel in TAE alongside an equivalent amount of undigested DNA to ensure the DNA was completely digested and products were the predicted size. The band containing the released transgene was

then excised from the gel and the agarose digested by treatment with agarase (Roche) according to the manufacturer's instructions.

The DNA was purified and concentrated using microcon 30 columns (Amicon) according to the manufacturer's instructions. Briefly, the DNA was passed through a microcon column and washed three times using 0.1mM EDTA/1mM Tris pH7.4. DNA was eluted in 10μl 0.1mM EDTA/1mM Tris pH7.4. The eluate was diluted 1:10 in transgenic buffer (0.1mM EDTA/10mM Tris pH7.4). The DNA concentration was determined by electrophoresis. DNA was stored at -20°C until the day of microinjection, at which point it was diluted in transgenic buffer to a final concentration of 2ng/μl and spun through a Spinex 0.22μm column (Costar).

2.5.4.2 Microinjection of recombinant DNA and oviductal transfers

The microinjections and subsequent oviductal transfers were performed by Lorna Purdie as described in MacKenzie *et al.*, 1997. Embryos used for microinjection were derived from [CBA x C57BL/6] F1 matings.

2.5.5 Identification of transgenic mice

All offspring from recipient female mice were ear clipped at weaning (21 days of age) for genotyping purposes in order to identify those in which the recombinant DNA construct had integrated. Genotyping was as described in section 2.5.1. Mice identified as transgenic constituted the founders of subsequent lines which were crossed with CD1 mice and subsequent generations genotyped to determine whether the transgene was faithfully transmitted.

2.5.6 Analysis of transgenic mice

2.5.6.1 *LacZ* staining protocol

Embryos were dissected in PBS and fixed in 4% PFA at 4°C for between 45 minutes (E8.5) and 3 hours (E13.5-E16.5) depending upon age. Following fixation, embryos were first rinsed in PBS and then washed in detergent wash 3 times for 20 minutes each at room temperature. Embryos were stained overnight in Xgal staining solution, protected from light, at room temperature. Ear tissue from ROSA26 mice, in which the *LacZ* gene is ubiquitously expressed, constituted positive controls. The following morning the staining solution was removed and embryos were rinsed in detergent wash twice for 20 minutes at room temperature before being fixed again in 4% PFA at 4°C overnight. The fix was removed and tissues were washed, and stored, in PBS.

2.5.6.2 Solutions

Detergent wash:

Compound	Per litre	Final concentration
Na_2HPO_4 (Fisher)	9.94g	0.1M, pH7.3
NaH_2PO_4 (BDH)	4.14g	0.1M, pH7.3
$MgCl_2$ (Riedel-de Haen)	0.406g	2mM
Sodium deoxycholate (Sigma)	1g	0.1%
Nonidet P-40 (ICN Biomedicals Inc.)	200μl	0.02%
BSA (Sigma)	0.5g	0.05%

Xgal staining solution:

Compound	Per 250ml	Final concentration
NaCl (BDH) (5M)	3.6ml	0.085%
$K_3Fe(CN)_6$ (Sigma)	410mg	5mM
$K_4Fe(CN)_6$ (Sigma)	525mg	5mM

The solution was made up to 250ml with detergent wash.

Prior to use, 30μl Xgal substrate (50mg/ml in DMF) was added per 5ml staining solution required.

2.6 Detection of gene expression

2.6.1 RNA *in situ* hybridisation

2.6.1.1 Preparation of labelled riboprobes

RNA products were kept on ice as much as possible and gloves were worn at all times to prevent exposure to degradative RNase enzymes. Electrophoresis equipment was soaked in 0.1M NaOH and rinsed in DEPC-dH_2O prior to use to destroy RNases.

Either linearised plasmid DNA containing cloned gene fragments, or PCR products amplified from cDNA, constituted templates for *in vitro* synthesis of labelled antisense RNA probes (Table 2.2). Plasmid DNA was cleaved at the 5' end of the insert with a suitable restriction endonuclease, whilst the RNA polymerase promoter used for transcription flanked the 3' end of the insert. PCR-generated template DNA was amplified using a primer pair for which theT7 promoter was incorporated into the reverse primer. In each case, the DNA was purified by gel extraction. Template DNA was transcribed using the appropriate RNA polymerase (T3, T7 or SP6) and nucleotides labelled with DIG (digoxigenin) using a DIG RNA labelling kit (Roche) according to the manufacturer's instructions. Template DNA was digested by the addition of 2μl RNase-free DNaseI for 45 minutes at 37°C. Purification of the riboprobe was completed using NucTrap Probe Purification columns (Stratagene) according to the manufacturer's instructions. 2μl RNase inhibitor (Roche) was added to the probe before storage at -20°C. The success of the DIG-labelling reaction was assessed by electrophoresis using a 1% gel run at 120V for 10 minutes.

2.6.1.2 Wholemount *in situ* hybridisation

Embryos aged between E9.5 and E15.5 were processed for wholemount *in situ* hybridisation. Embryos were dissected in DEPC-PBS and the extraembryonic membranes saved for genotyping purposes. In cases where the gut tissue alone was subject to *in situ* hybridisation, foreguts were dissected from embryos.

2.6.1.2.1 Solutions for wholemount *in situ* hybridisation:

In order to protect the RNA from degradation, solutions were made using DEPC-dH_2O, DEPC-PBS and DEPC-SSC for all steps prior to, and inclusive of, hybridisation.

dPBT

PBT was freshly prepared each time by the addition of 0.5ml 10% Triton X-100 (Sigma) to 50ml DEPC-PBS (final concentration 0.1%).

4% paraformaldehyde (PFA)

4% PFA (Sigma) was freshly prepared in DEPC-PBS.

Pre-hybridisation solution (prehyb)

Compound	Amount added per 50ml	Final concentration
De-ionised formamide (Sigma)	25ml	50%
20XDEPC-SSC	12.5ml	5X
Blocking reagent (Roche)	1g	2%
Triton X-100 (10%) (Sigma)	0.5ml	0.1%
CHAPS (10%) (Sigma)	2.5ml	0.5%
Yeast RNA (50mg/ml) (Sigma)	1ml	1mg/ml
EDTA (0.5M)	0.5ml	5mM
Heparin (10mg/ml) (Sigma)	250μl	50μg/ml

The solution was made up to 50ml with DEPC-dH_2O and all components were dissolved by gentle mixing.

Post-hybridisation solution (posthyb)

Compound	Amount added per 50ml	Final concentration
De-ionised formamide	25ml	50%
20XSSC	12.5ml	5X
Triton X-100 (10%)	0.5ml	0.1%
CHAPS (10%)	2.5ml	0.5%

The solution was made up to 50ml with dH_2O.

TNT

Compound	Amount added per 50ml	Final concentration
TRIS (1M, pH7.5)	2.5ml	50mM
NaCl (5M)	1.5ml	150mM
Triton X-100 (10%)	0.5ml	0.1%

The solution was made up to 50ml with dH_2O.

Blocking solution

Compound	Amount added per 50ml	Final concentration
Sheep serum*	7.5ml	15%
BSA	1g	2%
TRIS (1M, pH7.5)	2.5ml	50mM
NaCl (5M)	1.5ml	150mM
Triton X-100 (10%)	0.5ml	0.1%

*Previously heat inactivated at 50°C for 1 hour. The solution was made up to 50ml with dH_2O.

NMT

Compound	Amount added per 50ml	Final concentration
TRIS (1M, pH9.5)	5ml	0.1M
NaCl (5M)	1ml	0.1M
$MgCl_2$ (1M)	2.5ml	50mM

The solution was made up to 50ml with dH_2O.

2.6.1.2.2 Procedure for wholemount *in situ* hybridisation:

Following dissection of whole embryos/foreguts, tissue was fixed in 4% PFA at 4°C overnight. PFA was removed the following morning and the tissue washed with dPBT twice for 10 minutes each. Tissue not required for immediate use was dehydrated by washing for 20 minutes in increasing concentrations of methanol (MeOH): 25%, 50%, 75% MeOH in dPBT, followed by 3 washes in 100% MeOH before storage at -20°C. Prior to *in situ* hybridisation, dehydrated tissue was first rehydrated through a MeOH series consisting of decreasing concentrations of MeOH in dPBT: 75%, 50% and 25%. Whole embryos or dissected guts were subsequently washed 3 times in dPBT for 5 minutes each prior to treatment at room temperature with 10μg/ml proteinase K (Roche) dissolved in dPBT. The treatment time varied according to the size of the tissue: E10.5, E12.5 and E14.5 foreguts were treated for 10, 20 and 30 minutes respectively and treatment times for younger or older tissue were decreased or increased accordingly; E9.5 and E10.5 whole embryos were treated for 20 and 25 minutes respectively. Tissue was then washed in dPBT 3 times for 5 minutes each prior to re-fixation in 4% PFA for 45 minutes on ice and subsequently a further 3 dPBT washes. Tissue was transferred to 2ml cryotubes containing prehyb solution and left at room temperature to equilibrate. A further change of the prehyb solution followed, with tissue again being allowed to equilibrate at room temperature. Tissue was then incubated in fresh prehyb at 65°C for 1 hour, after which time the solution was again changed and tissue returned to 65°C for a further 3-4 hours. To create the hybridisation solution 5μl RNA probe (section 2.6.1.1) was added to 100μl prehyb and the probe denatured by heating to 80°C for 3 minutes. Embryos/foreguts were hybridised overnight at 65°C in the hybridisation solution.

The following day, the probe was removed by a series of washes in posthyb solution, each for 10 minutes at 65°C: 100% posthyb, 75% posthyb/25% 2XSSC, 50% posthyb/50% 2XSSC, 25% posthyb/75% 2XSSC. The tissue was then washed twice in 2XSSC/0.1% CHAPS for 30 minutes at 65°C and twice in 0.2XSSC/0.1% CHAPS, also at 65°C for 30 minutes. After 3 washes in fresh TNT solution for 5 minutes each at room temperature, the TNT was replaced with blocking solution

(pre-filtered using a 0.45μm syringe filter; Millipore) for at least 4 hours (with 3 changes) at 4°C. Tissue was incubated at 4°C overnight, gently rocking, in anti-DIG-AP fab fragment (Roche) at a 1:2000 dilution in blocking solution.

The next day, the unbound anti-DIG antibody was removed by washing tissue 5 times in TNT/0.1% BSA (pre-filtered using a 0.45μm syringe filter) for 1 hour each at 4°C, gently rocking. After a further change, tissue was left in TNT/0.1% BSA at 4°C either overnight or up to 2 days.

2.6.1.2.3 Standard detection of riboprobes:

In preparation for staining, embryos/foreguts were washed 3 times for 20 minutes each in NMT with 0.1% Triton X-100 at room temperature, then 3 times for 10 minutes each in NMT. To visualise the signal, staining solution was prepared by the addition of 3.3μl/ml NBT (100mg/ml in DMF) (Roche) and 3.5μl/ml BCIP (50mg/ml in DMF) (Roche) to NMT. Tissue was stained in the dark and monitored until the stain developed, which usually occurred within 1 to 3 hours, depending upon the probe. The staining reaction was terminated by rinsing tissue twice in PBS on ice, before fixation in 4% PFA overnight at 4°C.

2.6.1.2.4 Fast Red detection of riboprobes:

In preparation for Fast Red staining, embryos/foreguts were washed 3 times in 0.1M TRIS pH8.2 for 20 minutes each. The staining solution was prepared by the addition of 1 Fast Red tablet (Roche) per 2ml 0.1M TRIS pH8.2, followed by vortexing the solution until the tablet was dissolved. Tissue was stained at 37°C in the dark, monitored and the reaction terminated as described in section 2.6.1.2.3. Fast Red detection permitted both brightfield and fluorescence imaging of the resultant gene expression pattern.

2.6.2 Expression analysis via RT-PCR

2.6.2.1 RNA extraction

Extraction of RNA was performed prior to expression analysis of various sequences by reverse transcription (RT)-PCR. During this procedure, extreme care was taken to avoid contamination of samples with RNases. All solutions were made with DEPC-dH_2O, separate reagents reserved for RNA work only and surfaces of benches and glassware treated with commercially available RNase inactivating agents. Gloves were worn at all times and reagents kept on ice until required. RNA was extracted as quickly as possible after obtaining samples in order to minimise the degradation of crude RNA by limiting the activity of endogenous RNases.

E10.5 C57BL/6 embryos were harvested and heads and bodies separated before the tissue was frozen immediately on dry ice and stored at -70°C. To extract total RNA, the tissue was removed to dry ice and subsequently homogenised in 1ml RNAzolB (Biogenesis). 100μl chloroform was then added and the sample vortexed for 15 seconds before being left on ice for 5 minutes. The sample was then centrifuged for 15 minutes at 13,000rpm, 4°C, and the upper layer removed to a fresh screw cap eppendorf tube. An equal volume of isopropanol was added, the sample vortexed for 15 seconds and then left on ice for 15 minutes. A further centrifugation was performed as before, the supernatant discarded and the pellet washed in 70% EtOH (made up in DEPC-dH_2O). The DNA was re-suspended in 50μl DEPC-dH_2O. The concentration of the RNA was determined by spectrophotometry as described in section 2.1.4.

2.6.2.2 cDNA synthesis

A First Strand cDNA Synthesis kit (Roche) was used according to the manufacturer's instructions for cDNA synthesis prior to RT-PCR. PCR primers are detailed in Table 2.1 and PCR reactions were performed as described in section 2.3.

2.7 Analysis of gene expression

Gene expression patterns were analysed in both whole, and sectioned, tissue. Tissue was sectioned using either a vibratome, a microtome, or virtually following processing of tissue for OPT analysis. Each technique required different embedding procedures.

2.7.1 Agarose embedding

Fixed tissue was embedded in agarose prior to vibratome sectioning. The tissue was first incubated at 65°C in increasing concentrations of LMP agarose (Invitrogen) made up in dH_2O: 1%, 3% and 5%, being allowed at least 30 minutes to sink at each stage. The molten 5% LMP agarose containing the tissue was then poured into a suitably sized mould or small petri dish, orientated under a microscope using a pair of forceps and monitored at room temperature whilst the agarose set. When set, the agarose block containing the tissue was trimmed with a surgical blade (Swann Morton) and glued to the vibratome mount in the desired orientation for sectioning.

2.7.1.1 Vibratome sectioning

The agarose-embedded tissue was sectioned at 70-150μm on a vibratome. Each section was carefully lifted from the vibratome waterbath onto a superfrost slide (BDH). A drop of glycerol aqueous mount (Gelatine, 10g; Glycerol, 70ml; dH_2O, 60ml; stored at 50°C) was added to the slide and sections were then covered by a glass coverslip (thickness no.1) (BDH). To further prevent the drying of sections by evaporation, and to protect the tissue, the edges of the coverslip were sealed using vulcanising solution (Rema Tip Top, Germany). The mounted sections were stored in the dark at 4°C.

2.7.2 Wax embedding

To facilitate close histological analysis of thin sections of foregut tissue, the guts were embedded in wax prior to microtome sectioning. Prior to wax embedding, guts were dehydrated in 100% EtOH, then washed for 30 minutes in Xylene 3 times before they were transferred to molten wax. In order to ensure full penetration, the molten wax was changed 3 times with each incubation being performed for 1 hour at 65°C. Following transferral to a suitable mould, the wax containing the specimen was allowed to set at room temperature, during which time the tissue was orientated using a pair of forceps. The wax block was trimmed to the desired size and the edges angled to improve the capture of sections.

2.7.2.1 Microtome sectioning

The wax-embedded gut tissue was sectioned at 5μm on a microtome and strings of sections floated out in a 42°C waterbath. With the aid of a fine-tipped paintbrush, the sections were transferred to superfrost slides and the slides incubated at 65°C overnight to allow sections to dry. To remove the wax, slides were washed 3 times in Xylene for 10 minutes each, followed by 3x10 minute washes in 100% EtOH. Slides were then passed through a graded EtOH series, washed in dH_2O and haematoxylin and eosin staining performed according to a standard protocol. Tissue sections were mounted in DePeX (BDH) mounting medium.

2.7.3 Optical projection tomography (OPT) analysis

2.7.3.1 Preparation of non-fluorescence-stained tissue

Fixed tissue was embedded in 1% LMP agarose made up in sterile dH_2O. When set, the agarose block was trimmed with a surgical blade, ensuring adequate agarose was left surrounding the tissue, and glued to a cylindrical metal mount in the desired orientation for scanning. The samples were then each placed into a clean glass

container (height 5.5cm) and tissue dehydrated at room temperature overnight by the careful addition of 100% MeOH to cover the sample.

The following morning, tissue was cleared by replacing the MeOH with fresh BABB (1 part Benzyl Alcohol (Sigma) to 2 parts Benzyl Benzoate (Sigma)). Samples were monitored every 15-30 minutes and scanned by OPT when the tissue reached the desired transparency.

2.7.3.2 Preparation of fluorescence-stained tissue

Fast Red stain (section 2.6.1.2.4) is leached out of tissue by MeOH and BABB treatment, thus fixed tissue was instead cleared by passage through a glycerol series from 10% glycerol/dH_2O, increasing in 10% increments, to 80% glycerol/dH_2O at room temperature. Tissue was embedded in 2% LMP agarose made up in 80% glycerol/dH_2O and, when set, the agarose block was trimmed and mounted as described in section 2.7.3.1. The samples were then stored in a glass container containing 80% glycerol/dH_2O in preparation for scanning.

2.7.3.3 OPT scanning and data analysis

Samples were scanned in a cuvette filled with either BABB or 80% glycerol by both transmission and fluorescence OPT on an OPT scanner developed by Dr. James Sharpe at the MRC Human Genetics Unit, Edinburgh. For further information regarding OPT protocols, please refer to http://genex.hgu.mrc.ac.uk/OPT_Microscopy/. Data reconstructions were performed and 3D rotating movies were produced using scripts written by Dr. James Sharpe. GIMP software (a Unix version of Photoshop) was used to capture still images.

2.7.4 Microscopy

Tissue sections were analysed using a Zeiss Axioplan II fluorescence microscope (Carl Zeiss Ltd., Welwyn Garden City, UK) equipped with colour additive filters (Andover Corp, Salem, NH) for sequential colour imaging and a Photometrics CoolSnap HQ monochrome CCD camera (Roper Scientific, Tucson, Arizona). Image capture was by scripts written for IPLab Spectrum (Scanalytics Inc., Fairfax, VA) which controlled camera capture and filter selection via motorised filter wheels (Ludl Electronic Products Ltd., Hawthorne, NY).

Confocal microscopy, used to visualise Fast Red-stained tissue, was performed with the aid of a Zeiss LSM510 confocal system attached to a Zeiss Axiovert 100m microscope. The system was fitted with an Argon laser (488 and 514 lines), and two HeNe lasers (lines at 543nm and 633nm). Images were captured and viewed using Zeiss LSM510 v3.00 software.

2.8 Bioinformatics

The bioinformatics programmes and resources referred to in this thesis are listed below, together with the relevant world wide web (www) link or appropriate author.

Ensembl	http://www.ensembl.org
NCBI	http://www.ncbi.nlm.nih.gov
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
VISTA	http://www-gsd.lbl.gov/Vista (Mayor *et al.*, 2000)
PipMaker	http://bio.cse.psu.edu/pipmaker/ (Schwartz *et al.*, 2000)
NIX	available via http://www.hgmp.mrc.ac.uk
RepeatMasker	available via http://ftp.genome.washington.edu/ (Smit, A. and Green, E.; unpublished)
Emboss*	download via http://www.hgmp.mrc.ac.uk/Software/EMBOSS
CLUSTALW	available via http://www.hgmp.mrc.ac.uk (Higgins *et al.*, 1996)
Mega2 (version 2.1)	download via http://www.megasoftware.net/ (Kumar *et al.*, 2001)

* Used for sequence manipulation

Template	PCR conditions		Primer information	
	Anneal (°C)	Size (bp)	Direction	Primer sequence 5' => 3'
Bapx1	65	wt: 280 -/-: 1.4kb	Forward	CCGAACCAGAACAGCCGTGG
			Reverse	CAGCCCCCTTCCTGGAGAAC
DistB1	60-55*	1.7kb	Forward	GATCGGTACCCTTGTCGTCTTGAATCTTGC
			Reverse	GATCGGTACCTCTCTTGAAGACATGGCGAC
HPRT	55	295	Forward	CTGTAGATTTTATCAGACTGAAGA
			Reverse	GTCAAGGGCATATCCAACAACAAA
LacZ	58	400	Forward	TCTAGCCTGCAGGTCGAGGAC
			Reverse	GTGAGCGAGTAACAACCCGTC
Novel gene (RT-PCR)	55	1023/1837 (spliced/non-spliced)	Forward	CTTAGGACCTATTGCTCTGC
			Reverse	TGAGGCTTAAGGCTTGTACC
ProxB	60-55*	2kb	Forward	GATCGGTACCAGGACTTCAAGCTTAGACTC
			Reverse	GATCGGTACCGAAGATCTGTGATGTGTGTC
Rab28 (RT-PCR)	55	179/3002 (spliced/non-spliced)	Forward	GACTTGGAGCATATGCGAAC
			Reverse	TGTGACTGTTCTATTTCTGC
T3 primer*	55	1.7kb/2kb	-	AATTAACCCTCACTAAAGGG

Table 2.1: A list of PCR primers used, including the annealing temperature used in the PCR programme and the expected size of the PCR product. * Touchdown PCR programme (section 2.3.2). * Used in conjunction with DistB1 and ProxB primers.

in situ probe	Template	Restriction enzyme	RNA polymerase	Source
Bapx1	Plasmid	*Hind*III	T7	Lettice, L. A.
Barx1	Plasmid	*Bam*HI	T7	Hecksher-Sørensen, J.
Bmp2	Plasmid	*Xba*I	T3	Hogan, B.
Bmp4	Plasmid	*Acc*I	T7	Hogan, B.
Bmp7	Plasmid	*Sal*I	Sp6	Arkell, R.
Hox11	Plasmid	*Hind*III	T7	Hecksher-Sørensen, J.
Shh	Plasmid	*Hind*III	T3	McMahon, A.
Six2	Plasmid	*Eco*RI	T7	Oliver *et al.*, 1995
		Anneal (°C)	Primer sequence 5' => 3'	
Bmp11	PCR product	55	F: GAGCTGCGCCTAGAGAGCATCAAGTCG R: TAATACGACTCACTATAGGGGTTCTCTAGGACTCGAAGCTCCAT	Wright, A. J.
Nkx2.5	PCR product	65	F: ACTTGAACACCGTGCAGAGTCC R: TAATACGACTCACTATAGGGGTGTGGAATCCGTCGAAAGTGC	Hecksher-Sørensen, J.
Novel gene	PCR product	55	F: TCCAGGTACAAGCCTTAAGC R: TAATACGACTCACTATAGGGAGCTTCCATCGGTCTATGAC	Wright, A. J.
Rab28	PCR product	55	F: TCGGACACCATGTCAGACTC R: TAATACGACTCACTATAGGGCCCAATGATCCAAGGAGATC	Wright, A. J.

Table 2.2: A table detailing the origin of the *in situ* probes. The restriction enzyme cleaving at the 5' end of the insert is shown together with the appropriate RNA polymerase for production of an antisense probe. The T7 promoter sequence incorporated into reverse primers is indicated in red. The annealing temperature used in the PCR programme is included for each primer pair. F, forward; R, reverse.

Chapter 3 Attempts to locate mesenchymal regulatory elements

3 Attempts to locate mesenchymal regulatory elements

3.1 The recombineering approach

3.1.1 Introduction

Defining the precise role of the mesenchyme, delineating critical time points for its presence and identifying the local factors required for specialised regional function, would contribute greatly to our understanding of normal development. However, considering the diverse sites at which mesenchyme functions, the countless processes in which it is involved and the consequent immense genetic complexity, tackling these questions becomes an enormous task. Research in the past has therefore tended to focus on the more thorough study of a single organ or morphogenetic process (Bard, 1990). The current investigation was designed to probe gut development in particular, although it could be applicable to additional regions of the organism. The idea was to exploit the fact that *Bapx1* is expressed in a subset of mesenchymal cells from the early stages of embryogenesis. The restricted expression of *Bapx1* in the mesenchyme provides the means by which this specific cell population can be targeted using *Bapx1*-expressing sites in the embryo. It is not known whether the recessive lethality of the *Bapx1* knockout mutation (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999) is a direct consequence of an essential requirement for the mesenchyme. Nor have critical time points for mesenchyme presence or reasons for its potential indispensability been determined.

3.1.2 Experimental strategy

Conditional targeted ablation of specific cell populations in living transgenic animals is a powerful strategy to analyse cell functions *in vivo*. It has been successfully employed to generate mouse models of disease, to explore the role of particular tissue types during development and to conduct cell lineage analysis (Rindi *et al.*, 1999; Lee *et al.*, 2000; Saito *et al.*, 2001). Conditional targeted ablation is accomplished, in transgenic mice, by expressing a cytotoxic gene under the control of a tissue-specific

enhancer/promoter which in this case would derive from the *Bapx1* gene. The objective was to target an inducible toxic phenotype to *Bapx1*-expressing cells at specific developmental stages in order to investigate more precisely the importance of the mesenchyme to gut development. The making of the targeting construct would involve the recently developed and reportedly highly efficient recombineering system (discussed below).

Prior to the induction of cell ablation, complete and correct temporally and spatially regulated *Bapx1* expression as directed by the flanking sequence contained within the bacterial artificial chromosome (BAC) or P1 artificial chromosome (PAC) genomic insert must be verified. This would be accomplished by insertion (by recombination) of a GFP-*LacZ* reporter cassette into the *Bapx1* locus, followed by detailed analysis of expression domains. The creation of a reporter cassette would allow several lines of further study, such as deletion analysis of flanking sequence to identify *Bapx1* regulatory elements responsible for gut-specific expression. This in turn would provide a means to manipulate the gut mesenchyme, for example by the misexpression of genes, and has the potential to greatly advance our biological understanding of gut development.

Other potential applications of recombineering to investigate the role of *Bapx1* include the creation of *Bapx1-Cre* transgenic mice in which the gene encoding the Cre recombinase is targeted to the *Bapx1* locus and is thus directed to *Bapx1*-specific domains during development. Defined matings with other genetically-modified mice could be performed to activate, for example, a cytotoxic gene to accomplish tissue-specific cell ablation; a reporter gene to verify appropriate expression; or a *LaacZ* gene for cell lineage analysis. An alternative strategy would be to modify the *Bapx1* locus in embryonic stem (ES) cells however this is a highly time-consuming, expensive and technically difficult procedure, hence the decision to pursue the recombineering route.

3.1.3 Issues addressed by manipulation of the mesenchyme

There are a number of biologically important questions that such an experiment was envisioned to address. It was of particular interest to determine the effects of conditional cell ablation upon molecular and morphological boundaries within the gut and thus investigate the role of the mesenchyme in gut development. Furthermore, the stage-specific phenotypes of conditional mutants could be compared to wildtype and *Bapx1* null embryos to shed light on the relative importance of *Bapx1*-expressing mesenchyme at particular stages of embryogenesis. Although the focus is gut development, *Bapx1* expression is much more widespread than the gut mesentery. The potential therefore exists to ascertain whether cell ablation would produce a limb phenotype (unlike the *Bapx1* knockout), in addition to examining the effect on the axial skeleton and craniofacial sites of *Bapx1* expression. Moreover, the evolutionary conservation of *Bapx1* and widespread involvement of the mesenchyme in inductive interactions during development renders the results potentially applicable to numerous developmental systems.

3.1.4 The advantages of recombineering

With the completion of finished sequence for the human genome, the next major challenge towards understanding our developmental programme will be to determine how each of our genes functions in isolation, and in the context of the genome. Elucidating the intricate control points and signalling pathways underlying the tight regulation of developmental gene expression in a single organ alone is an immense task. The undertaking of such complex functional genomics, which will rely on studies of model organisms, will require the manipulation of large segments of DNA. Recombinogenic engineering, or recombineering, is a novel form of chromosome engineering which became possible within the last few years (Murphy, 1998; Zhang *et al.*, 1998; Muyrers *et al.*, 1999; Yu *et al.*, 2000; reviewed by Copeland *et al.*, 2001). It was permitted by the development of highly efficient phage-based *Escherichia coli* homologous recombination systems. The latter enabled large segments of genomic DNA such as those carried on BACs or PACs to be modified and subcloned, without requiring restriction enzymes or DNA ligases. Such classical *E.coli*-based

recombinant DNA technology breaks down when cloning vehicles and the target site contain hundreds of kilobases of DNA. Hence the ability to *modify* BAC clones became the primary limiting factor following the advent of BAC cloning technology in *E.coli*.

The crucial advantage of recombineering is the ability to target and manipulate a specific genomic region using electroporated linear dsDNA or ssDNA (Yu *et al.*, 2000; Ellis *et al.*, 2001; Swaminathan *et al.*, 2001). Notably, only short (30-50bp) regions of flanking homology are required (Yu *et al.*, 2000) and selective markers, which may affect the function of the region being studied, need not be permanently added to the BAC.

Host DY380 cells were created by introducing a defective λ (lambda) prophage into DH10B cells, a BAC host strain (Lee *et al.*, 2001). The defective λ prophage supplies functions that protect and recombine a linear DNA targeting cassette with its substrate sequence by homologous recombination, eliminating the requirement for standard cloning. Recombineering thus has the potential to simplify the generation of transgenic mouse models, extend the experimental spectrum and provide a more precise and refined genetic analysis of the genome.

3.1.5 Results: Inability to target the *Bapx1* locus

Recombineering experiments were undertaken using two *Bapx1*-containing PAC vectors. Unfortunately, despite a number of attempts to modify the *Bapx1* locus contained within the PACs, the recombineering experiments were unsuccessful. An inherent problem with recombineering technology appears to be the modification of *any* desired locus. The location of the target site within its genomic context together with the precise modification desired may be critical factors in modulating the success of the procedure. This could explain why, for some researchers, recombineering has proven successful whilst others, including both myself and colleagues within the Human Genetics Unit (personal communication; Pete Budd), have experienced limited success and ultimately been forced to abandon it in favour of other approaches. A brief resumé of the experimental specifications associated with the recombineering protocol is included in Appendix 1.

3.2 Comparative sequence analysis of the *bagpipe* region

3.2.1 Introduction

The inability to use recombineering to assess the developmental role of the mesenchyme necessitated the employment of an alternative strategy. Ultimately the aim remained the same: to manipulate the mesenchyme such that expression of selected genes could be effected in a *Bapx1*-specific manner, hence allowing for misexpression experiments or mesenchyme ablation. Defining the regulatory elements responsible for the *Bapx1* expression pattern constituted an alternative approach since it was envisaged that these elements could then be used to direct expression to mesenchymal cells.

Understanding the mechanisms that underlie gene regulation is one of the major challenges as we strive towards a more complete understanding of genome function. A better comprehension of these processes will establish a critical foundation on which to assimilate the collective complexities of developmental biology. *Bapx1* promoter activity has been located immediately upstream of the gene, occupying position –880 to –844 relative to the published 5' end of *Bapx1* exon1 (+1) (Rodrigo *et al.*, 2003). This region was found to contain binding sites for the paired-box transcription factors Pax1 and Pax9, which were proposed to act directly on *Bapx1*, which in turn initiated sclerotome chondrogenesis. Nevertheless, transgenic mice harbouring the p10kb vector (see Introduction) containing approximately 7kb of upstream sequence failed to exhibit correct *Bapx1* expression. These results suggest that although a certain degree of promoter activity may reside within the 5' 1kb, additional regulatory elements likely exist external to the 7kb 5' of the gene.

The search for regulatory elements is a formidable task; functional elements in non-coding DNA sequences may reside at varying distances from the target gene. They can be found downstream (Choi and Engel, 1986), within an intron (Kohler *et al.*, 1996) and, even when located upstream of their cognate gene, may act over enormous distances. For example, a long range enhancer of *Shh* which is separated from the gene by 1Mb has recently been reported (Lettice *et al.*, 2003). Clearly the cost and nature of transgenic analysis is generally prohibitive to the initiation of

bioassays without having first narrowed down somewhat the region in question. This assessment and elimination process is greatly facilitated by employing a bioinformatics approach.

Comparative sequence analysis is a powerful means of detecting functional regions in genomic sequences. The rationale behind this approach is the assumption that sequence conservation implies functional constraint. It has been well documented that during the course of evolution, functional parts of sequences tend to be more highly conserved than non-functional regions (Hardison *et al.*, 1997; Hardison, 2000; Wasserman *et al.*, 2000). This applies not only to genes, but also their associated regulatory elements and ever since the high degree of conservation of the latter was first recognised (Aparicio *et al.*, 1995), comparative DNA sequence analyses have become increasingly important.

Computational gene prediction has been facilitated by the prodigious amount of genomic data now available and it has become a routine process. Software, for example NIX (available via the HGMP website http://www.hgmp.mrc.ac.uk), integrates searches for such features as consensus splice signals, similarity to known genes or proteins, open reading frame (ORF) length and codon usage. Such information is combined to build a picture of the compositional landscape and ultimately to predict gene structures. By contrast, the unambiguous identification of *cis*-acting regulatory elements is far from straightforward and analogous *in silico* prediction programs do not exist. Thus the identification of conserved functional regions in non-coding DNA relies on comparisons of DNA sequences between distantly related genomes. Such an approach has proven successful in a number of studies, for example the identification of regulatory elements in vertebrate *Hox* gene clusters for which several known regulatory sequences were also recovered (Santini *et al.*, 2003).

'Phylogenetic footprinting' is a term used to describe the preferential conservation of functional sequences over the course of evolution by selectional pressure. Further to the identification of *regulatory sequences*, comparison of orthologous genomic regions will no doubt also play an important role in ascribing *functions* to these non-coding sequences (Pennacchio and Rubin, 2001). Knowledge of the identity of a mediating transcription factor can give important insights into the function of a gene via inference of the processes or conditions that lead to expression.

Significant research progress in the modelling of DNA binding sites (Bulyk *et al.*, 2001) has resulted in the binding profiles of an increasing number of transcription factors having been established and collated (JASPAR database: http://www.phylofoot.org/consite; Sandelin *et al.*, 2004). Accordingly, new bioinformatics tools are emerging that utilise transcription factor binding profiles and phylogenetic footprints to detect the target binding sites of various transcription factors in conserved non-coding DNA sequences (Lenhard *et al.*, 2003; Johansson *et al.*, 2003). Thus by integrating knowledge of regulatory elements with potential binding factors, it is possible to achieve a deeper understanding and a more comprehensive picture of the regulatory pathways within cells.

The optimal phylogenetic distance over which one can find regulatory elements remains unresolved. Humans and mice diverged approximately 90 million years (myr) ago (Hedges and Kumar, 2002) yet share a high degree of sequence conservation, as demonstrated by the partitioning of greater than 90% of the human and mouse genomes into corresponding regions of conserved synteny (Waterston *et al.*, 2002). One must therefore take into account the "signal-to-noise ratio"; comparisons may be more beneficial if they include more distantly related species.

Teleosts such as the pufferfish *Fugu rubripes* and the zebrafish *Danio rerio* are the most distant extant vertebrate precursors of mammals with an estimated separation time of 450 myr (Hedges and Kumar, 2002). Initial characterisation of the *Fugu* genome based on random sequencing and screening of a genomic library with single copy probes revealed a 400Mb genome containing a low percentage of repetitive DNA and small introns (Brenner *et al.*, 1993). The sequencing of *Fugu* has since confirmed a vertebrate genome size that is one eighth that of humans and in which gene density is predictably high (Aparicio *et al.*, 2002). This genome compaction, together with the great evolutionary distance, render it greatly advantageous to include *Fugu* in comparative sequence analyses. In particular, inclusion of *Fugu* sequence should increase the powers of detection of conserved putative regulatory elements. Although the sequencing and assembly of zebrafish is not so advanced, where sequence is available it is likely to provide similar benefits to those of *Fugu*. Furthermore, various properties of zebrafish such as their ease of manipulation, both physically and genetically, and initial transparency are recognised as conducive to comprehensive studies of development, in particular embryology.

Consequently they are widely used in the research community and the increased sequence availability, leading to the detection of putative *cis*-acting regulatory sequences, therefore has the potential to be used directly to corroborate other research. For example, by providing additional links between the genomes of different organisms and their associated transgenic studies.

3.2.2 Results

3.2.2.1 Bapx1 protein conservation

As an initial assessment of the evolutionary importance of *Bapx1*, the protein sequences from seven different species were obtained via a link from either the Ensembl genome browser (http://www.ensembl.org) or NCBI website (http://www.ncbi.nlm.nih.gov). *Drosophila* was the most phylogenetically diverged species included; estimates regarding the split between the arthropods and the lineage that gave rise to the vertebrates range from 833 to 993 myr ago (Durand, 2003). The resulting sequence alignment (Figure 3.1), produced using the CLUSTALW program (Higgins *et al.*, 1996) illustrates that a number of regions in the peptide are identical, including between species as distant as humans and fruitflies. *Bapx1* was first isolated as an evolutionarily conserved homologue of *Drosophila bagpipe* (*bap*) by 'zoo' Southern blot analysis (Tribioli *et al.*, 1997) and, being a homeobox gene, it is perhaps not surprising that the 60 amino acid homeodomain is highly similar between the seven species. Three other domains, with functions not yet defined, namely the TN domain (Harvey, 1996), NK-2SD domain (Harvey, 1996) and Bap domain (Newman *et al.*, 1997), are also remarkably well conserved (Tribioli *et al.*, 1997; Newman *et al.*, 1997) (see Figure 3.1).

Interestingly, a putative function of the TN domain has been revealed by the demonstration of sequence similarity within the TN domains of Nkx6.1 and Nkx2.2 to a core region of the engrailed homology-1 (eh1) domain present in *Drosophila* Engrailed (En), a transcriptional repressor (Muhr *et al.*, 2001). This conserved eh1 motif mediates recruitment of Gro/TLE corepressors, which are mammalian relatives of the prototypic Drosophila Gro/TLE protein Groucho (Gro), and it was

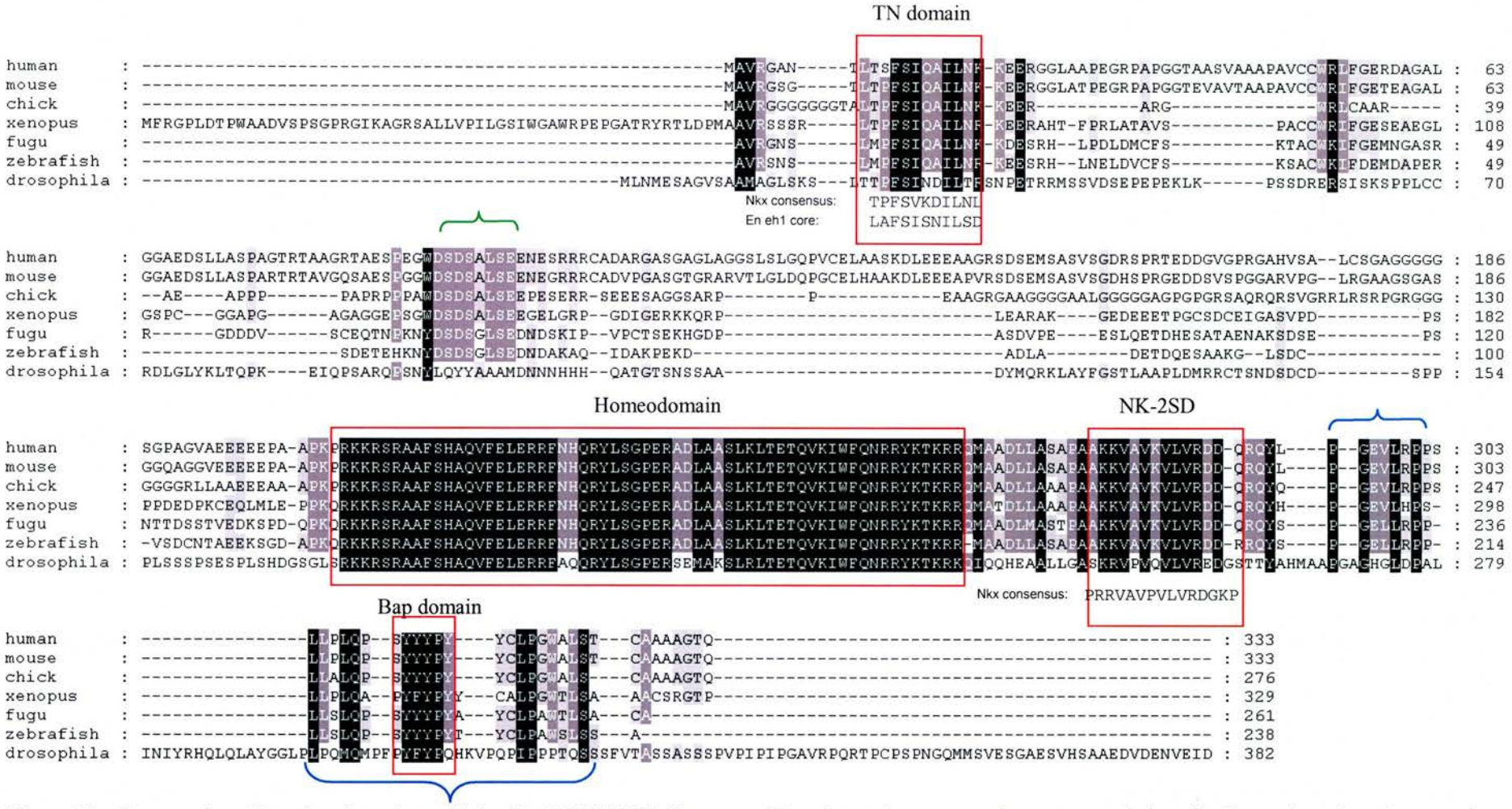


Figure 3.1: Conservation of Bapx1 at the polypeptide level. CLUSTALW alignment of Bapx1 protein sequences from seven evolutionarily diverged species. Conservation levels of amino acid similarity at aligned positions are denoted by shading as follows: black, 100%; dark grey, 80% or greater; light grey, 60% or greater; no shading, less than 60%. The highly conserved homeodomain, TN domain, NK-2SD domain and Bap domain are boxed and Nkx consensus sequences for the TN and NK-2SD domains are indicated. The core sequence of the Engrailed (En) eh1 domain is shown for comparison. Additional conserved regions are bracketed in blue and a vertebrate-specific region is bracketed in green. Amino acid numbering is indicated to the right.

demonstrated that Nkx6.1 and Nkx2.2 interact with Gro/TLE corepressors via their TN domains (Muhr *et al.*, 2001). Comparison of the Bapx1 TN domain to the Nkx consensus determined by Harvey reveals that seven of eleven amino acids are identical (Figure 3.1). Moreover, direct comparison of the Bapx1 TN domain to the core sequence of the eh1 domain present in En reveals the sharing of five amino acids (Figure 3.1), suggesting that Bapx1 (Nkx3.2) may also interact with mouse Gro/TLE corepressors.

In the case of Nkx6.1 interaction with Grg4, it was suggested that other domains may contribute to the interaction (Muhr *et al.*, 2001). It is noteworthy that the Bapx1 alignment reveals the existence of several additional well conserved amino acids at various positions in the Bapx1 protein, but generally clustered around the extreme N terminal or C terminal ends (Figure 3.1). These additional regions of conservation include one that appears to be vertebrate-specific. The reason for this conservation of specific residues is unknown but it is presumably due to their presence being critical for the folding and/or *in vivo* function of the peptide, the latter including further protein-protein interactions. Whether there also exists a vertebrate-specific domain is an intriguing prospect.

The available sequence for all NK3 family members currently determined was retrieved via the NCBI browser and aligned using CLUSTALW. In addition to Bapx1 (Nkx3.2) and the ancestral bap in *Drosophila*, these sequences included members of the Nkx3.1 and Nkx3.3 (to date only identified in amphibians) families. Mouse Nkx2.5 was included to root the tree. Mega2 software was subsequently employed to construct a Neighbour Joining tree and bootstrap the phylogeny using 1000 replications. The resulting phylogenetic tree (Figure 3.2) demonstrates the clustering of distinct NK3 families and supports current evolutionary models as regards the relatedness of the species. *Drosophila* bap occupies a separate branch to the other NK3 proteins whilst Nkx2.5 has diverged significantly from all of the NK3 proteins. Interestingly, although only *Nkx3.2* (*Bapx1*) homologues have been established to date in fish, there exists a zebrafish partial cDNA clone (Accession number: BQ449759; 496bp) which exhibits a high degree of sequence similarity with mammalian *Nkx3.1* sequence. Thus it appears that a complete cDNA may reveal the existence of a zebrafish *Nkx3.1* gene and, by extension, help to refine estimates of the evolutionary time point at which this family diverged from its *Nkx3.2* paralogues.

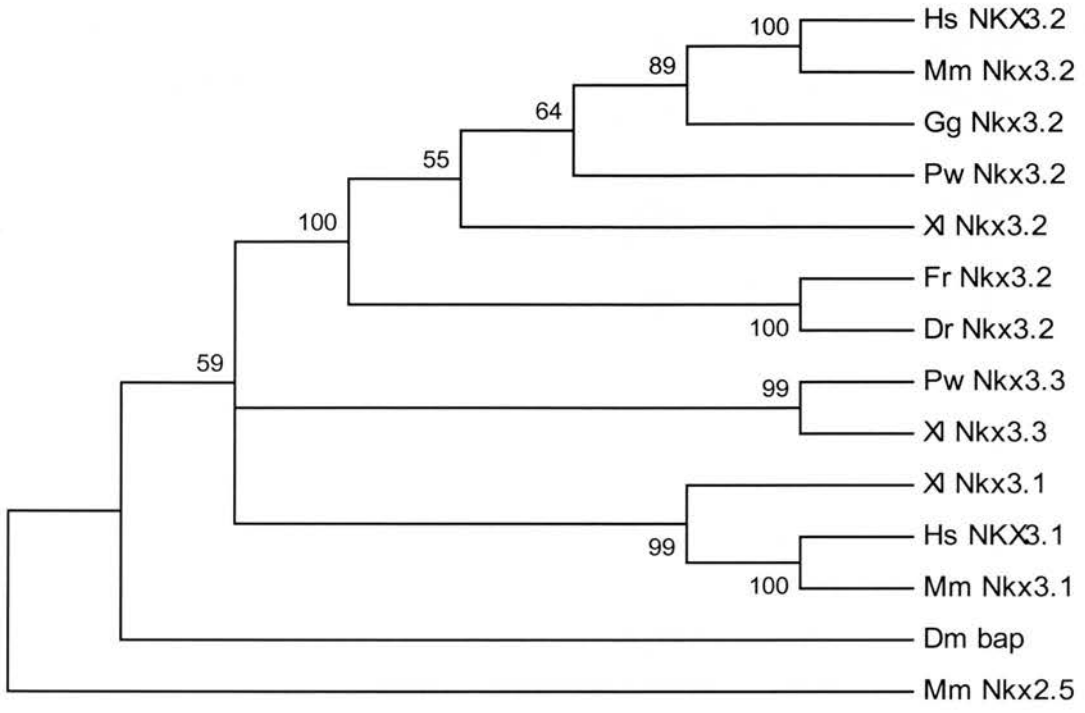


Figure 3.2: Evolutionary relatedness of NK3 proteins from eight evolutionarily diverged species. A phylogenetic tree produced using the Neighbour Joining method (with 1000 bootstrap replications) from a CLUSTALW alignment of NK3 protein sequences including those of Bapx1 (Nkx3.2) in Figure 3.1. The Nkx3.2, Nkx3.3 and Nkx3.1 proteins form distinct clusters within the tree and lie on a separate evolutionary branch to the single NK3 member in *Drosophila*, bap. Mouse Nkx2.5 forms the root of the phylogenetic tree. Numerical values given at branch points of the tree indicate the probability that this grouping is correct, where 100 denotes absolute certainty. Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Fr, *Fugu rubripes*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Pw, *Pleurodeles walt1*; Xl, *Xenopus laevis*. The *Xenopus* homologue of *Nkx3.3* is known as *zampogna* (*zax*) (Newman and Kreig, 1999) whilst the *Nkx3.1* homologue is named *koza* (Newman and Kreig, 2002).

Collectively the results substantiate that Bapx1 is a highly conserved protein and, although it appears to have acquired additional roles during vertebrate evolution supplementary to its function in the visceral mesoderm in *Drosophila*, certain regulatory elements are predicted to be shared across vertebrate species.

3.2.2.2 Conservation of synteny in the *Bapx1* region

The first step in the identification of putative *Bapx1* regulatory elements was to obtain genomic sequence from the species to be compared. Accordingly, the latest versions of the human (*Homo sapiens*; v.16.33.1), mouse (*Mus musculus*; v.16.30.1), *Fugu* (*Fugu rubripes*; v.16.2.1) and zebrafish (*Danio rerio*; v.16.2.1) assemblies were downloaded from the Ensembl browser. The availability of contiguous sequence for human and mouse has facilitated the alignment of syntenic genomic regions. Figure 3.3 shows that synteny is conserved between human chromosome 4p and mouse chromosome 5 in the 4.3Mb region surrounding *Bapx1*. At the outer ends of this region in both cases lies *HS3ST1* on one side and *C1QTNF7* on the other. The latter is inferred in the case of *Mus musculus* from homology matches between *C1QTNF7* and the "novel" gene. Mouse *2700023P08Rik* is a likely homologue of *RAB28*; *AF013969* shares significant homology with *Q9P2L9* and the middle of the three neighbouring "novel" genes to *AF013969* appears to be the mouse equivalent of *NM_148894* (see later). Although Ensembl predicts two other "novel" genes immediately flanking the *NM_148894* homologue, they may not be genuine and indeed have no sequence matches listed in other species.

The *Fugu* and zebrafish genomic sequence is not yet fully assembled and presently comprises multiple scaffolds or chromosome fragments respectively. Nevertheless, there exists a 116kb *Fugu* scaffold (scaffold_776) which contains the *Fugu* homologues of *RAB28* (*RB28_RAT*), *BAPX1* (*P70061*) and *NM_148894* (Figure 3.3). The latter is depicted as "novel" in Figure 3.3 yet Ensembl BLAST analysis demonstrates significant identity to *NM_148894*. Both this and the *RAB28* gene are recorded as having protein products whose functions are unknown. The corresponding *RAB28* ESTs derive from a multitude of different tissues suggesting that it is ubiquitously expressed (UniGene cluster Hs.404189). Similarly, there exist a number of different cDNA sources of *NM_148894* (UniGene cluster Hs.205442).

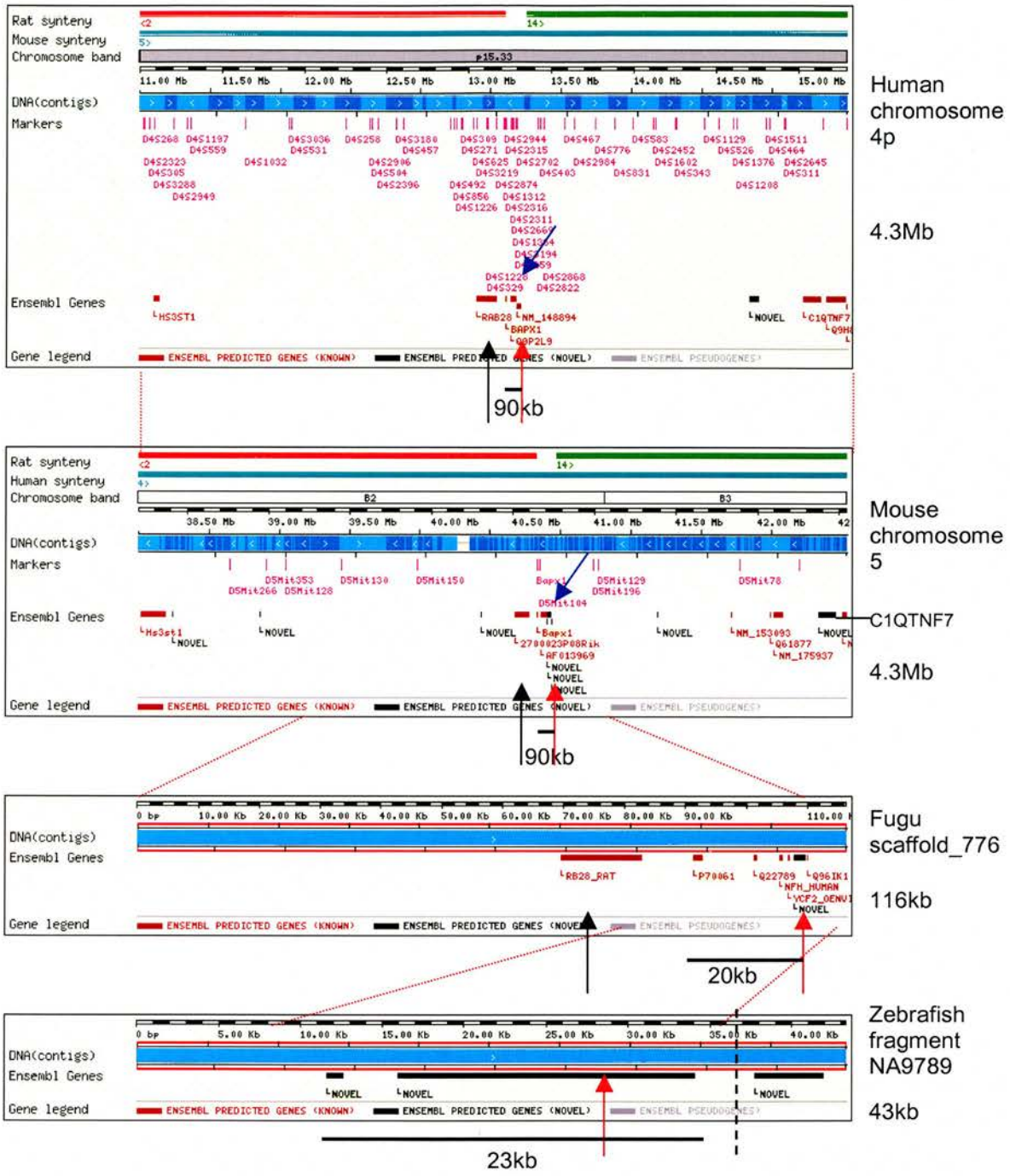


Figure 3.3: Conservation of synteny and genomic compaction in the *Bapx1* region. The *Bapx1*-containing regions of human, mouse, *Fugu* and zebrafish genomic DNA as viewed in Ensembl. The *Fugu* and zebrafish sequences are subregions of the human and mouse region, as denoted by red dotted lines. The black dotted line indicates a break in synteny between mammals and fish. The immediate *Bapx1* flanking region between *NM_148894* and *RAB28* is conserved between humans, mice and fish. Length of genomic sequence encompassed for each species is indicated on the right. The mouse homologue of *C1QTNF7* is indicated as "novel". Ensembl nomenclature for homologous genes varies between species: The black arrow points to *RAB28* homologous genes (presumably insufficient sequence for inclusion of the zebrafish equivalent), the red arrow indicates *NM_148894* and the blue arrow marks *Q9P2L9* and its homologue in mouse. *Bapx1* is highlighted in green. The length of a comparable region between the four species is shown in each case beneath a black horizontal line. Note that scale bars indicating sequence length do not begin at zero in the human and mouse sequence.

Interestingly, there does not appear to be a *Fugu* homologue of *Q9P2L9*. The latter is recorded as an antigen-containing epitope to monoclonal antibody MMS-85/12, its function is unknown and it appears to be ubiquitously expressed (UniGene cluster Mm.154431). Perhaps this gene is superfluous to the *Fugu* genome, permitting considerable local sequence divergence. Although Ensembl predicts five genes in close proximity to the 3' end of scaffold_776 (Figure 3.3), only the "novel" gene is substantiated by a match to a gene in another species (*NM_148894*). Furthermore, the structure of the "novel" gene is more convincing, consisting of five exons to produce a 361 amino acid protein. The other four predictions all comprise a single exon only and may not be *bona fide* genes. Indeed sequence alignments demonstrate that exons of *NM_148894* and *BAPX1* are generally well conserved between human, mouse and *Fugu* whilst the intervening sequence appears to be considerably diverged in *Fugu* (see later).

Chromosome fragment NA9789 in zebrafish contains three "novel" genes within 43kb (Figure 3.3). The first, residing between 11kb and 13kb, hits *Bapx1* in a BLAST query whilst the second, covering approximately 18kb, identifies *NM_148894* as a close match. The third gene contains sequence matching that of a gene (*STK12*) on human chromosome 17. Pending confirmation of homology, this disruption of synteny is suggestive of a rearrangement in the ancestral sequence such that *NM_148894* and *STK12* now reside on different human chromosomes (4 and 17 respectively).

Collectively, the above results provide strong evidence for the conservation of synteny between humans and teleosts in the *Bapx1* region extending from *NM_148894* to *RAB28*. (It is deduced that the absence of a *RAB28* homologue in zebrafish fragment NA9789 results from insufficient sequence 5' to the *Bapx1* match; see Figure 3.3). The maintenance of synteny in a region having undergone 450 myr divergence is a good indication that functional constraint is operating to modify the rate of evolution. Consequently it is reasonable to believe that a closer inspection of the *Bapx1* flanking sequence may reveal important, tissue-specifc *Bapx1* regulatory elements.

3.2.2.2.1 Genomic compaction in *Fugu* and zebrafish

Genomic sequence from the start of *NM_148894* to the end of *RAB28* covers approximately 270kb in human and mouse. The equivalent region in *Fugu* occupies only 41kb, representing greater than a six fold compaction. In order to include zebrafish in the comparison, sequence is required that is available for all four species. Genomic sequence between and inclusive of *NM_148894* and *Bapx1* fulfils this criterion. In both human and mouse the region spans 90kb, in *Fugu* only 20kb and in zebrafish 23kb (Figure 3.3). Thus *Fugu* and zebrafish exhibit a marked, and similar, reduction in genome size; the *Bapx1* region occupies approximately 25% of the equivalent sector in mammals.

3.2.2.3 The search for putative *Bapx1* regulatory elements

Human, mouse and *Fugu* genomic sequences between and inclusive of *NM_148894* and *RAB28* were exported from Ensembl. The comparable portion of this region in zebrafish, contained on chromosome fragment NA9789, was also downloaded. Annotation files detailing exon positions for each of the genes were compiled using the sim4 program which compares cDNAs to large genomic sequences (Florea *et al.*, 1998). Preliminary searches were restricted to DNA sequences conserved between human and mouse as this can greatly enrich for regulatory motifs. Strong conservation alone can be sufficient to predict regulatory elements from pairwise alignments (Loots *et al.*, 2000). Various software is available that computes the alignment of similar regions in two or more DNA sequences; VISTA (Visualising global DNA sequence alignments of arbitrary length) (Mayor *et al.*, 2000) and PipMaker (Percent identity plot Maker) (Schwartz *et al.*, 2000) are two such programs. VISTA is accessible via http://www-gsd.lbl.gov/Vista and PipMaker via http://bio.cse.psu.edu/pipmaker/. VISTA was originally intended for the comparison of more closely related species and thus was chosen for the initial human versus mouse pairwise alignment. The resulting plot is shown in Figure 3.4.

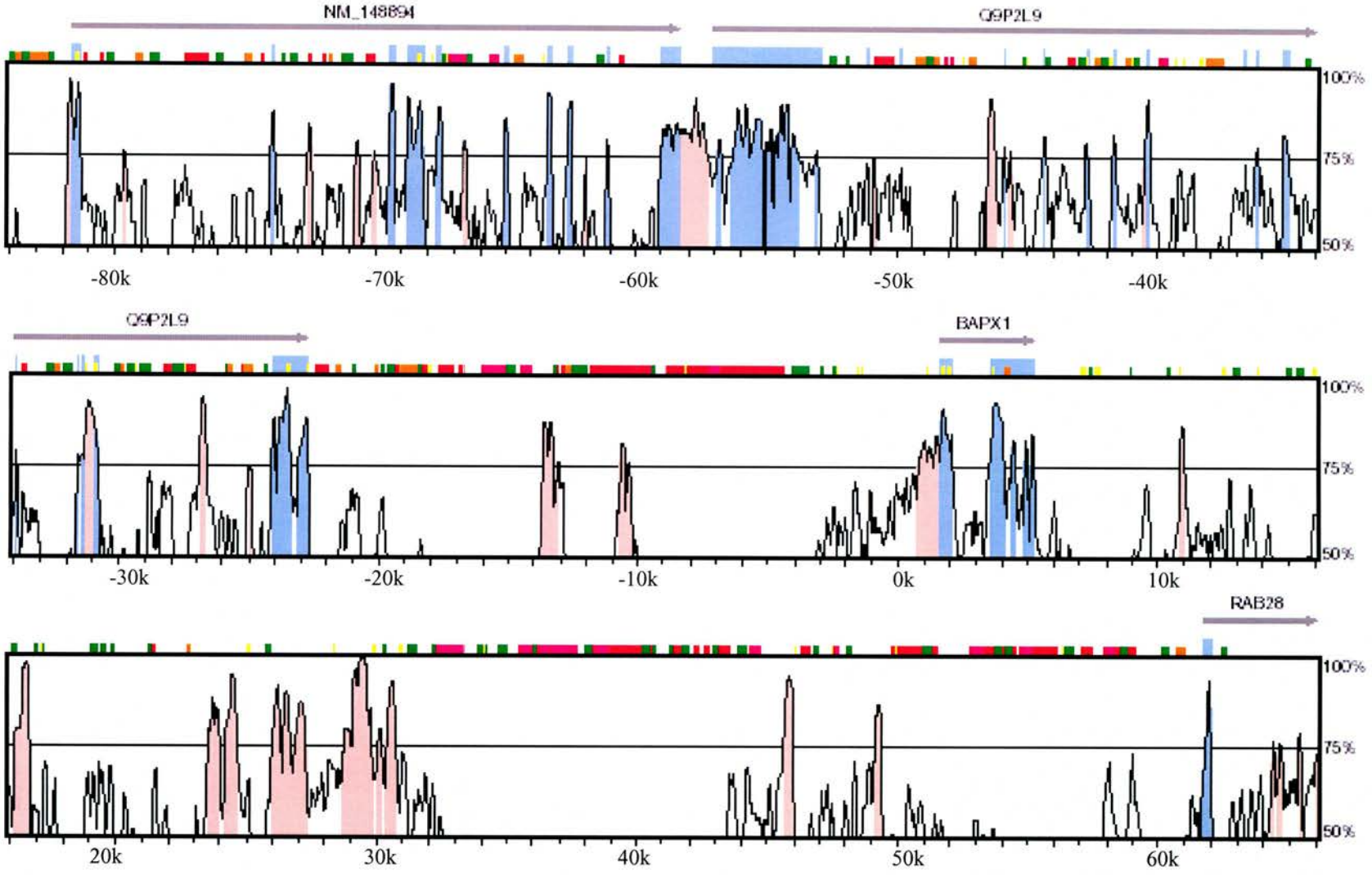


Figure 3.4: See following page for figure legend.

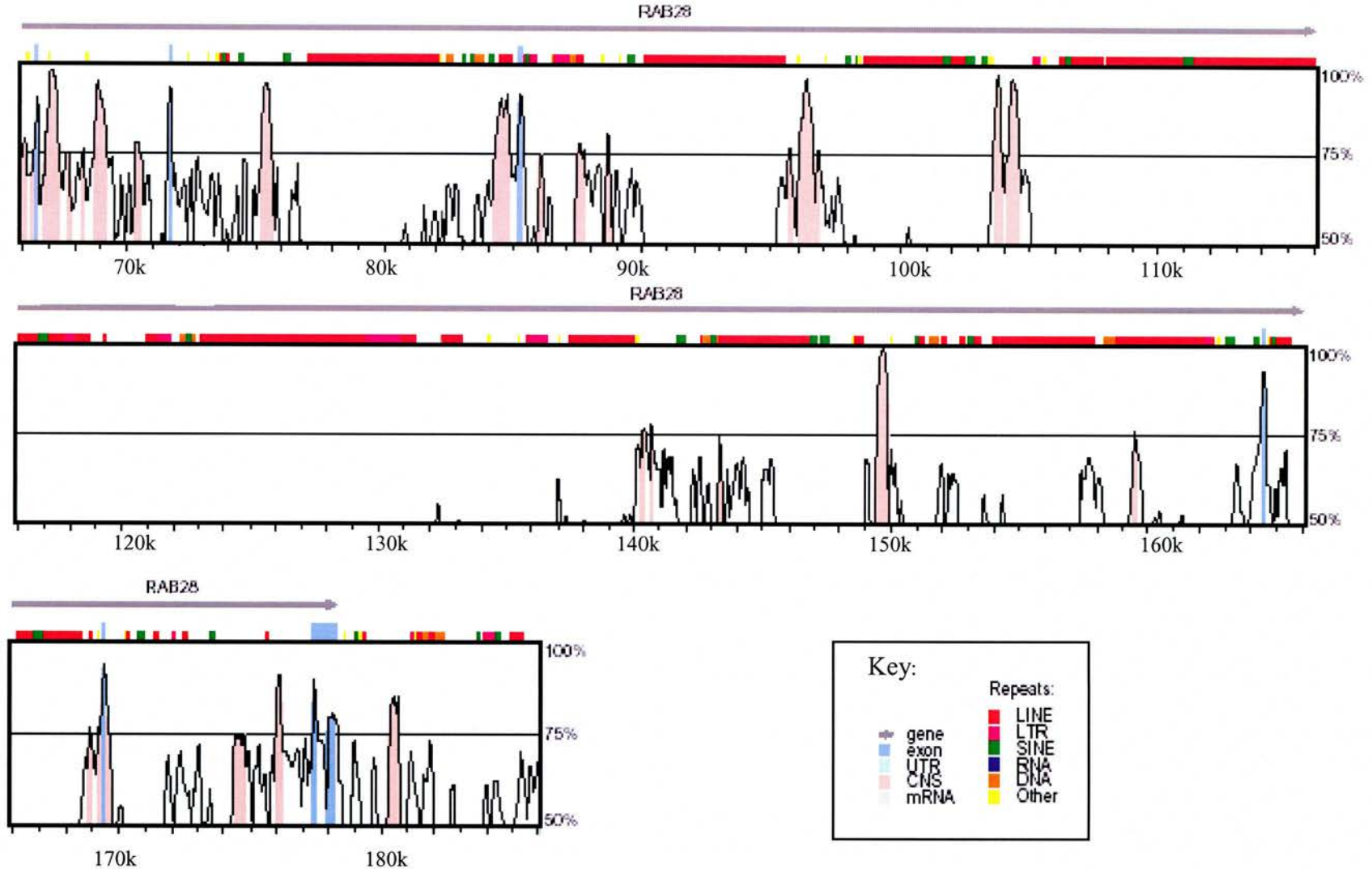


Figure 3.4: Human/mouse conservation at the nucleotide level. A pairwise VISTA alignment of human and mouse genomic sequence between and inclusive of *NM_148894* and *RAB28*. The base sequence is human and runs along the X-axis; relative distance is indicated by the scale bar and the approximate start of *BAPX1* is designated 0k. The Y-axis shows the extent of conservation above the minimum level set at 50% identity over 100bp. The direction of transcription is indicated by an arrow beneath the gene name. Exons appear as blue boxes above the graph and shaded blue regions within it. Conserved non-coding sequence (CNS) is shaded pink if it meets the minimum criteria of 75% nucleotide identity over at least 100bp.

3.2.2.3.1 Narrowing down the region of interest

The exons of all genes contained within this region are generally well conserved (greater than 75% nucleotide identity) between human and mouse. Interestingly, conservation upstream of *Bapx1*, where one would expect to find promoter elements, diminishes very rapidly and is only greater than 75% within 1kb 5' of exon1 (Figure 3.4). Notably this includes the proposed Pax1/Pax9 binding sites at position –880 to –844 (Rodrigo *et al.*, 2003). The intervening region between the last exon of *Q9P2L9* and the Pax1/Pax9 binding site 5' of *Bapx1* is essentially devoid of conserved regions except at two separate loci, both between 700bp and 1kb in length, located 12kb and 15kb upstream of *Bapx1*. Their conservation level is approximately 70-85%. Nothing of significance appears to lie in the 1.3kb *Bapx1* intron however a number of conserved regions appear to reside 3' of *Bapx1* (Figure 3.4). Two regions in particular are conspicuous as a result of their extremely high level of nucleotide identity in one case (peaking at almost 100%) and high level conservation more-or-less maintained over 8kb in the other (eight conserved regions ranging between 75% and 85% nucleotide identity). These potential regulatory elements reside 11kb and 18kb downstream of the 3' end of *Bapx1* respectively. Two additional, short conserved regions (86% and 79%) are found before the start of *RAB28*. The *Bapx1-RAB28* stretch looked most promising for the discovery of putative regulatory elements thus a more extensive search was conducted using this sequence alone.

3.2.2.3.2 Identification of two putative enhancer elements

A human/mouse pairwise alignment of the 56kb intervening sequence between *Bapx1* and *RAB28* identified the most prominent conserved sequences as those located 11kb and 18kb downstream of *Bapx1*, henceforth referred to as ProxB (for Proximal to *Bapx1*) and DistB (for Distal to *Bapx1*) respectively (Figure 3.5). To further investigate the importance of these putative regulatory elements, the *Bapx1* regions from *Fugu* and zebrafish were introduced into the comparison. The premise was that by increasing the phylogenetic distance encompassed by the comparison the power of detection of a functional CNS (conserved non-coding sequence) should improve.

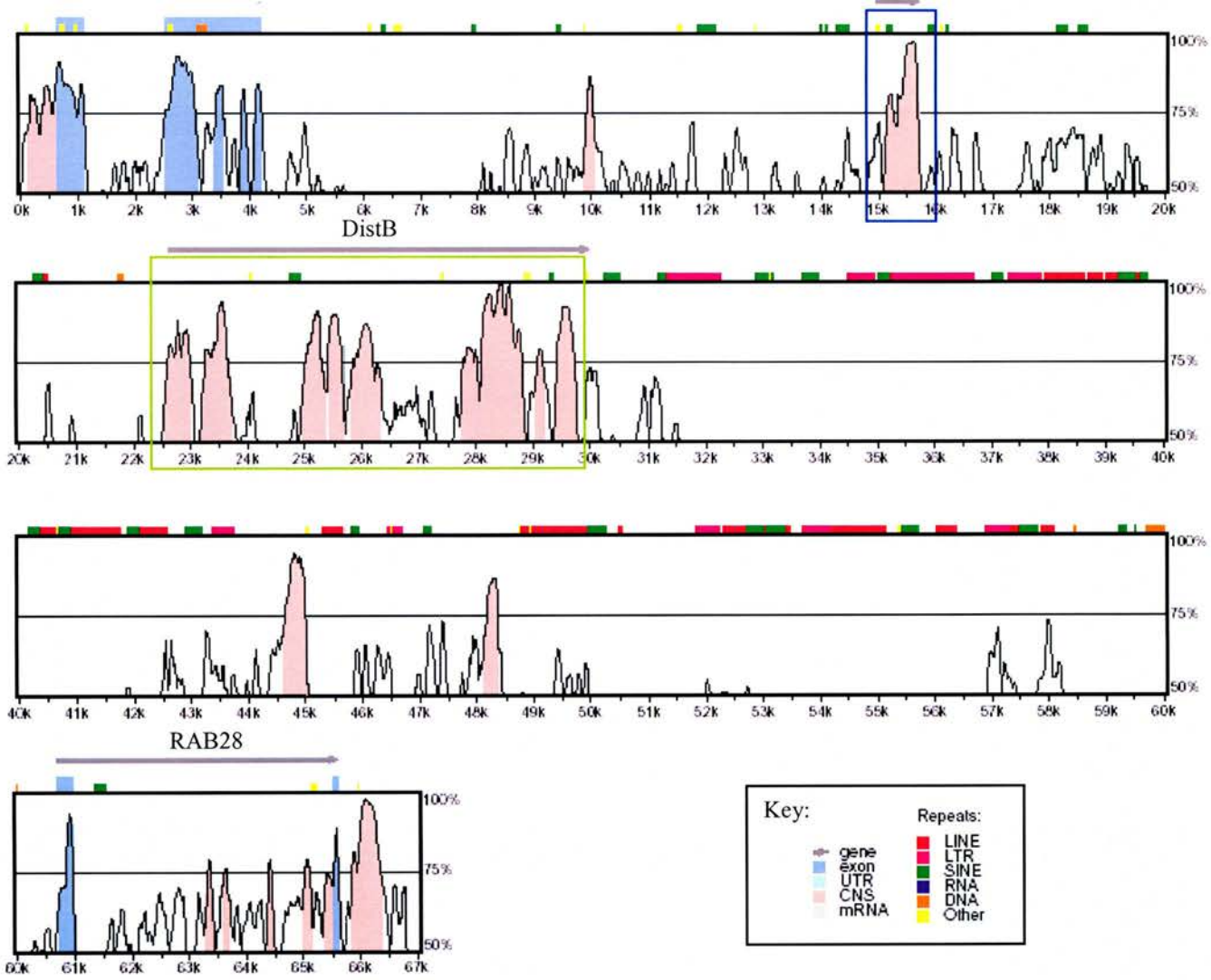


Figure 3.5: The identification of two putative *Bapx1* regulatory elements. A pairwise VISTA alignment of human and mouse genomic sequence running from *Bapx1* to the second intron of *RAB28*. The base sequence is human and runs along the X-axis; relative distance is indicated by the scale bar. The Y-axis shows the extent of conservation above the minimum level set at 50% identity over 100bp. Exons appear as blue boxes above the graph and shaded blue regions within it. Conserved non-coding sequence (CNS) is shaded pink if it meets the minimum criteria of 75% nucleotide identity over at least 100bp. Two putative *Bapx1* regulatory elements, termed ProxB (for proximal to *Bapx1*) and DistB (for distal to *Bapx1*), are boxed.

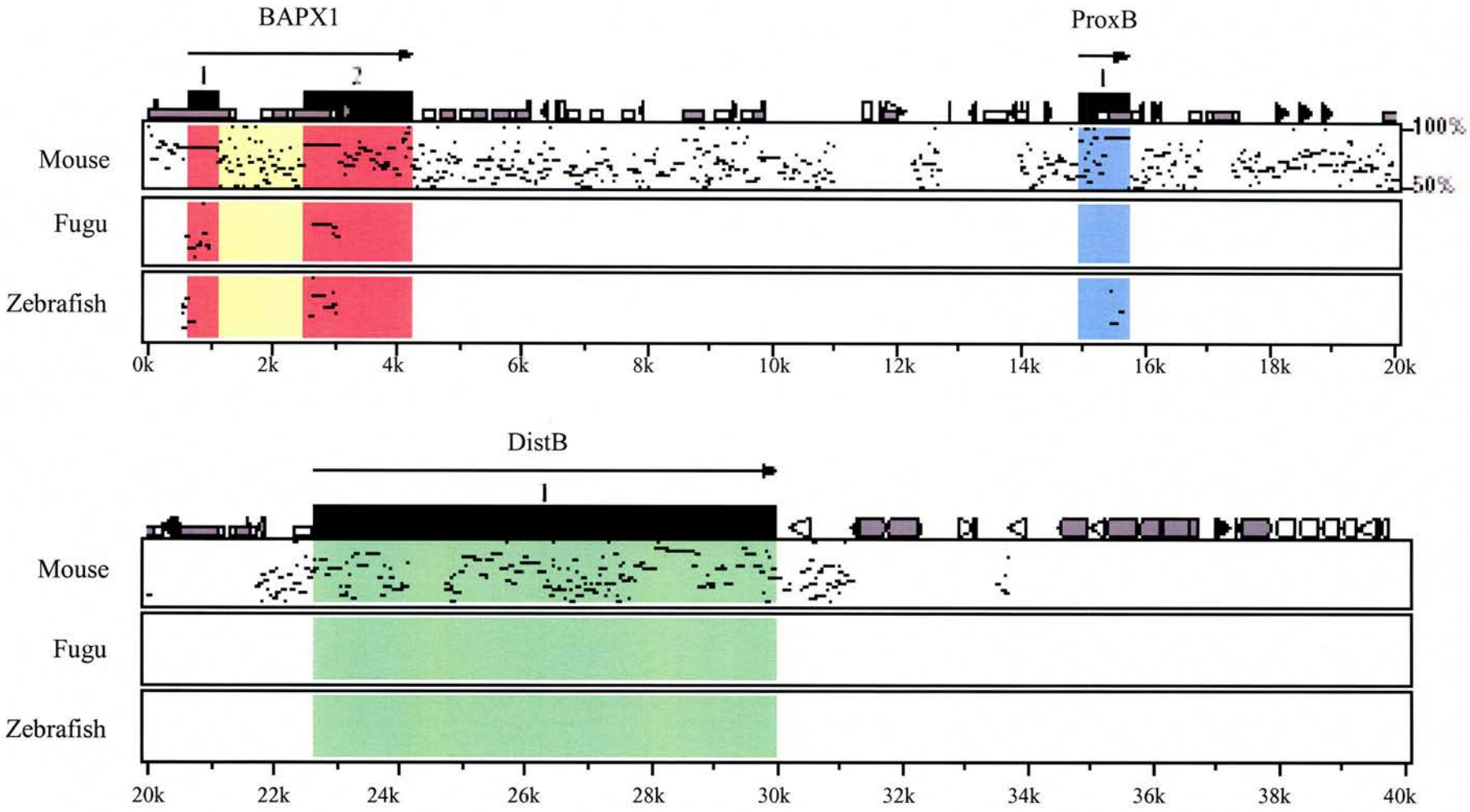


Figure 3.6: See following page for figure legend.

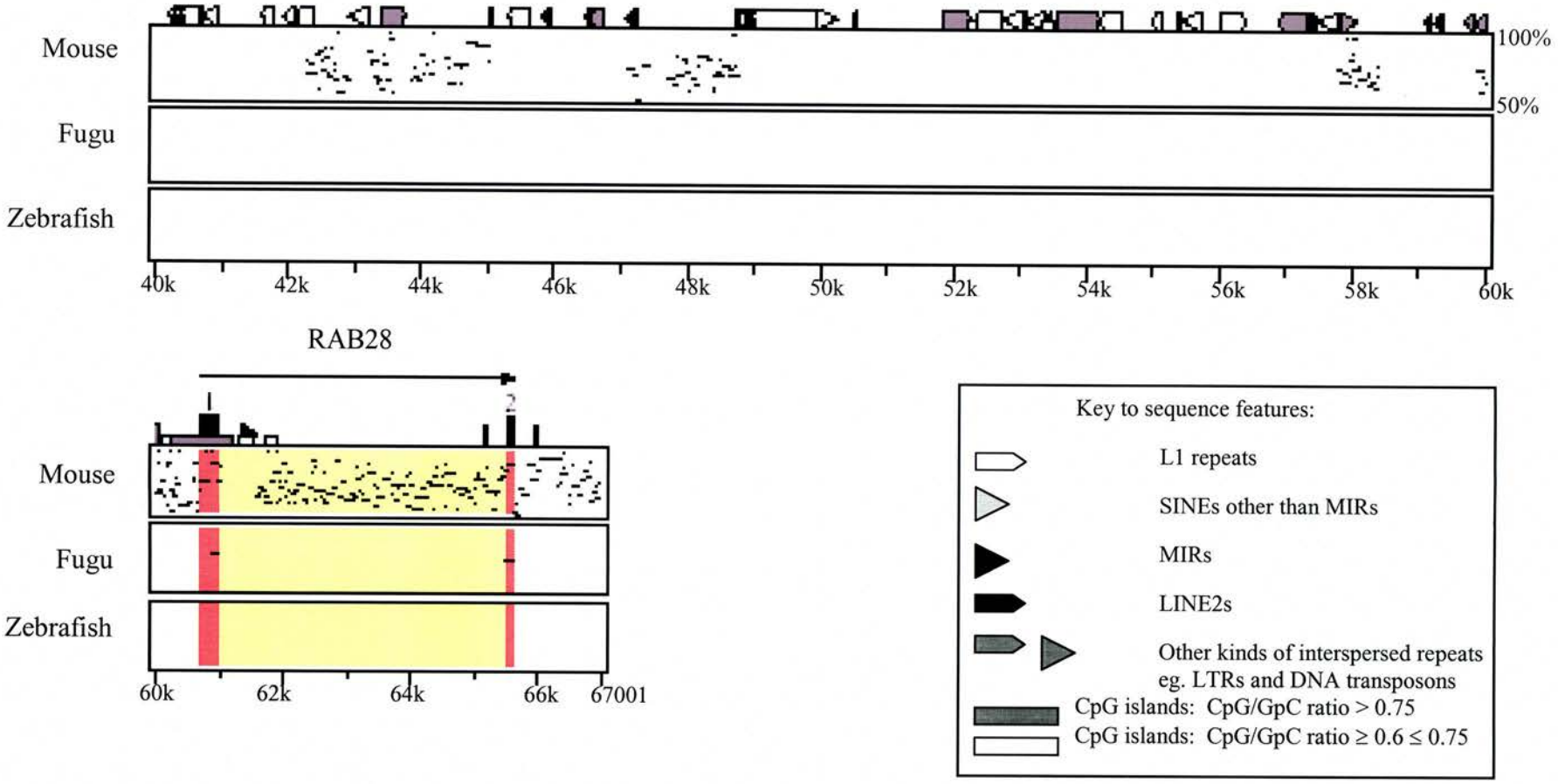


Figure 3.6: The conservation of ProxB to zebrafish. A multi-species Pip plot showing the alignment of the *Bapx1-RAB28* region in human, mouse, *Fugu* and zebrafish. Each species is assigned a separate row. The scale bar along the lower X-axis refers to the human base sequence, for which features are plotted along the upper X-axis. The length of the horizontal black line indicates the length of the conserved sequence in the given species whilst its position on the Y-axis indicates the extent of conservation above a minimum level of 50%. Genes are shaded yellow except for exons which are red and also appear as black boxes along the upper X-axis. Arrows above genes show the direction of transcription. The putative control element ProxB is shaded blue whilst DistB is shaded green. *RAB28* sequence is not present in zebrafish chromosome fragment NA9789. LINE, long interspersed nuclear element; LTR, long terminal repeat; MIR, mammalian interspersed repeat; SINE, short interspersed nuclear element.

MultiPipMaker was used to conduct a four way sequence alignment between human, mouse, *Fugu* and zebrafish *Bapx1-RAB28*. Figure 3.6 illustrates the resulting Pip plot. Both exons of *Bapx1* show conservation to both *Fugu* and zebrafish. Only the first two exons of *RAB28* were included in the analysis and both are conserved in *Fugu*. Their apparent absence from zebrafish results from there being insufficient sequence present for analysis 3' of *Bapx1* on fragment NA9789. The larger putative regulatory region, DistB, would appear to lack conservation to either *Fugu* or zebrafish DNA (Figure 3.6). However, the ProxB putative control region is conserved in zebrafish, significantly increasing the prospect of its being an evolutionarily conserved regulatory element. Moreover, the conserved region in both mouse and zebrafish corresponds to the location of a CpG island in the human ProxB sequence where the ratio CpG/GpC exceeds 0.75 (Figure 3.6). Notably, such islands are also located within both exons of *Bapx1* where conservation to *Fugu* and zebrafish is evident. Besides the two *Bapx1* exons, ProxB appears to be the only conserved sequence maintained in teleosts within the aligned genomic region. The portion of zebrafish fragment NA9789 showing homology with mouse ProxB is depicted in Figure 3.7.

A pairwise BLAST alignment of mouse ProxB and zebrafish ProxB was conducted using the BLAST program at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). Analysis of the result at the nucleotide level revealed conservation of 180bp at 68.33%. This figure was virtually identical when human and zebrafish ProxB were aligned. A repeat alignment of mouse and human ProxB, limited to the region conserved in zebrafish, revealed 96.67% nucleotide identity over 180bp. This corresponds to the peak at nearly 100% previously seen in Figures 3.4 and 3.5.

An additional Pip plot between human, mouse and *Fugu* was conducted for the entire *NM_148894-RAB28* region (Figure 3.8). This revealed three short sequences conserved to *Fugu* (between 63bp and 198bp in length), besides those corresponding to exons, which all fell within the fourth intron of *RAB28* (see Figure 3.8). These were also detected by VISTA which calculated conservation levels ranging from 67%-87% for human-*Fugu* and 66%-74% for mouse-*Fugu* alignments. *RAB28* is a large gene, occupying 117kb in human and mouse, and intron4 is by far the longest of the introns. The *Fugu* equivalent covers only 13kb, one ninth of the

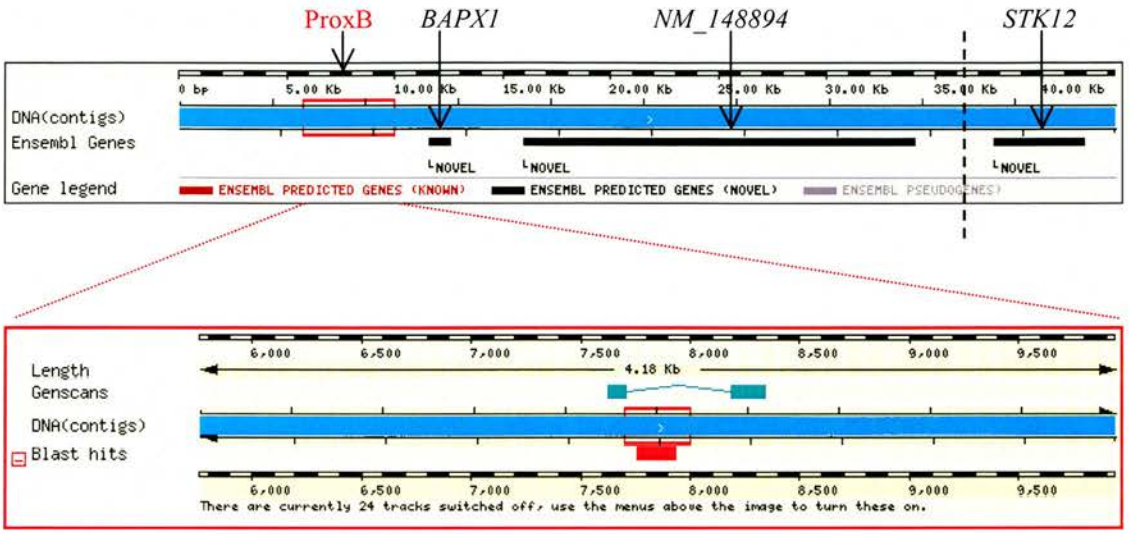


Figure 3.7: The location of ProxB in zebrafish. The upper portion depicts zebrafish chromosome fragment NA9789 as seen in Ensembl. Gene names are indicated and the conserved ProxB sequence is labelled and boxed in red. The black dotted line indicates a break in synteny between mammals and fish. The lower portion zooms in on the ProxB-containing region. The green linked boxes either side of ProxB mark an unsubstantiated gene prediction by the Genscan program. The BLAST hit identified by Ensembl is to the equivalent human and mouse ProxB sequences.

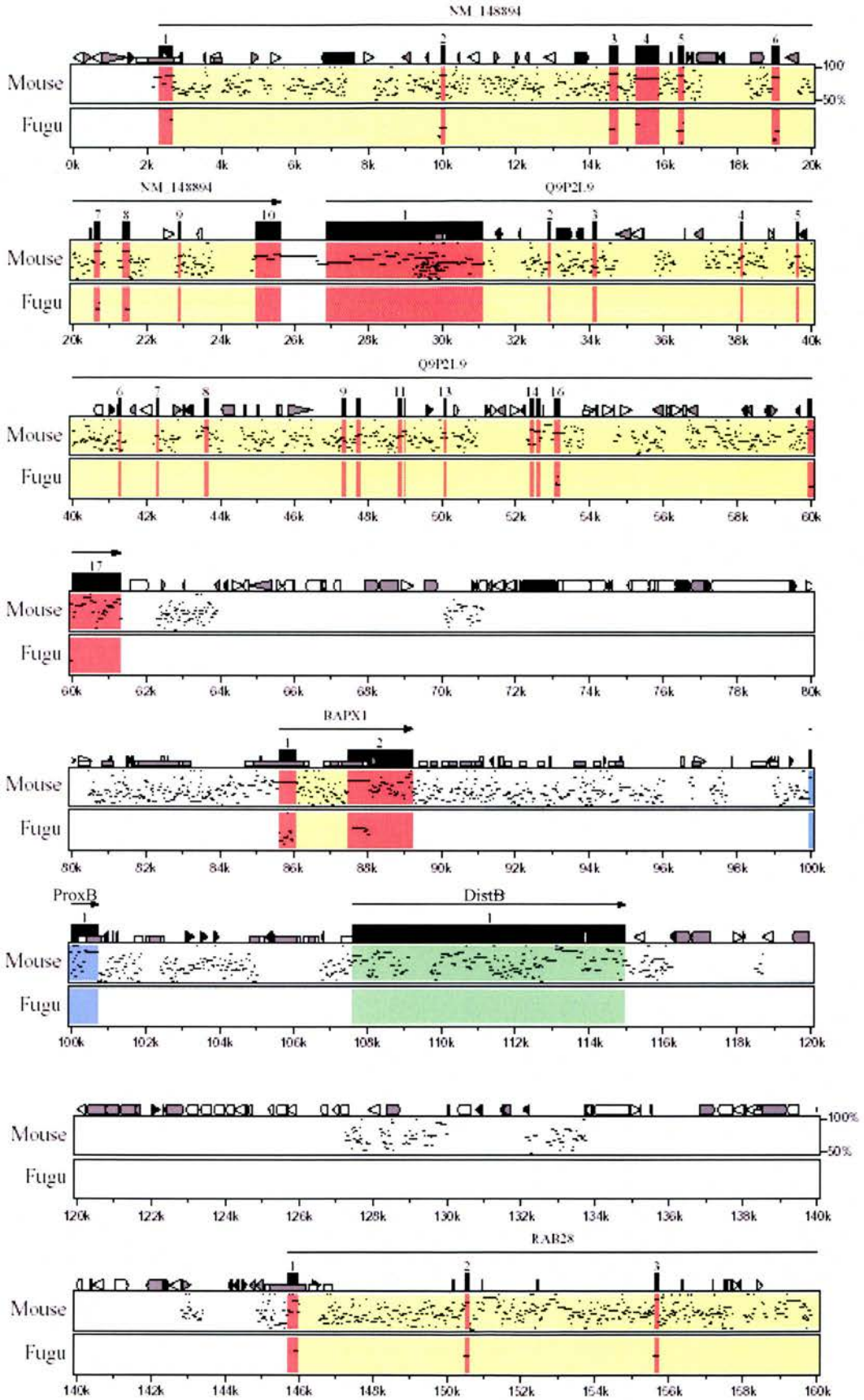
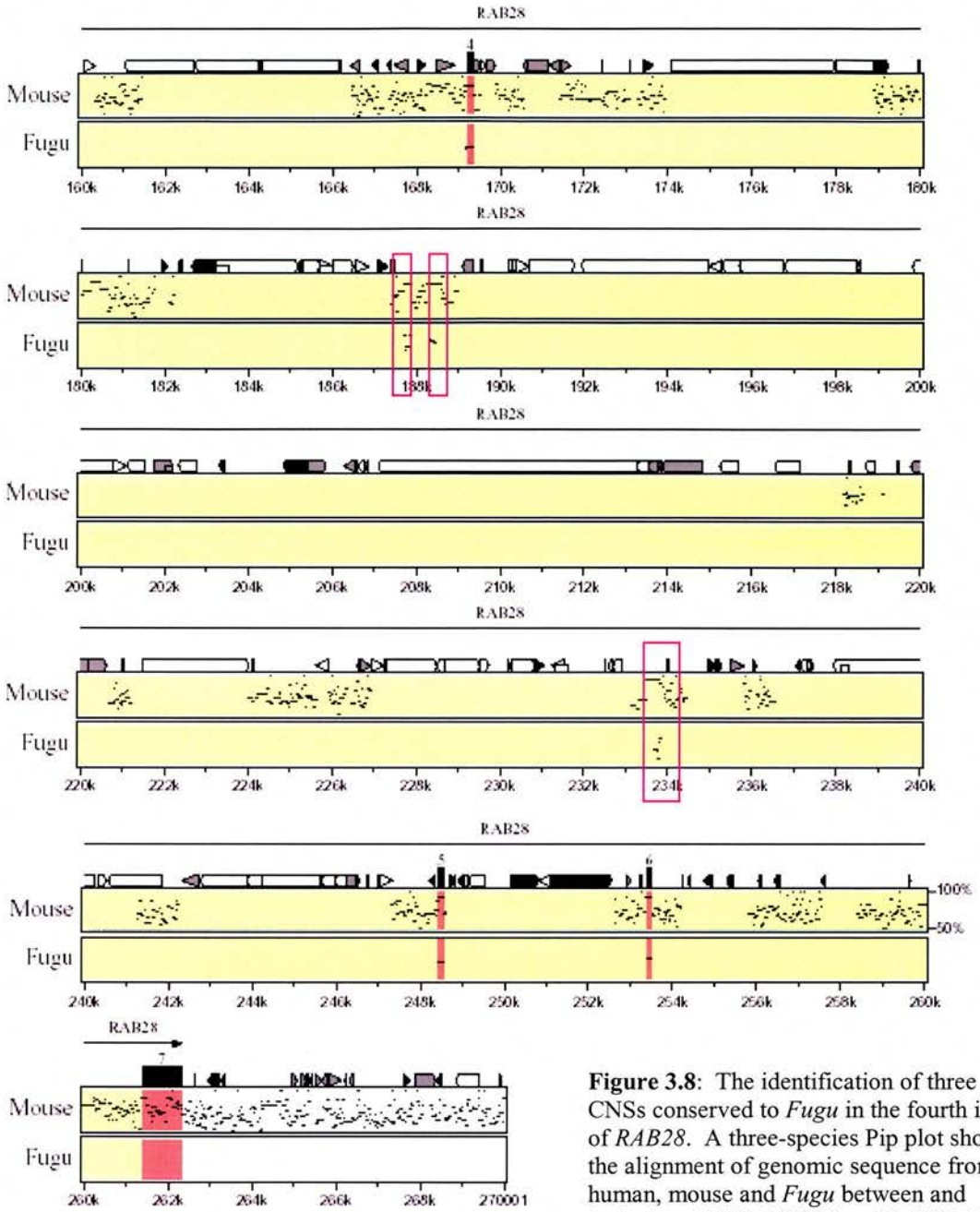


Figure 3.8: See following page for figure legend.



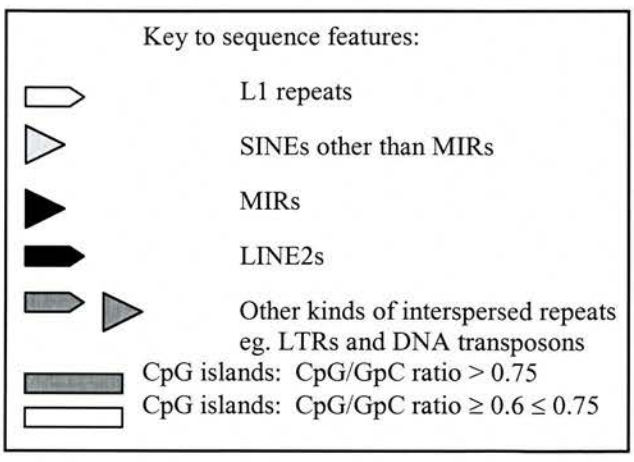


Figure 3.8: The identification of three CNSs conserved to *Fugu* in the fourth intron of *RAB28*. A three-species Pip plot showing the alignment of genomic sequence from human, mouse and *Fugu* between and inclusive of *NM_148894* and *RAB28*. The scale bar along the lower X-axis refers to the human base sequence, for which features are plotted along the upper X-axis. Sequences conserved to mouse and *Fugu* appear as short, black horizontal lines in their corresponding row. The length of the horizontal black line indicates the length of the conserved sequence in the given species whilst its position on the Y-axis indicates the extent of conservation above a minimum level of 50%. Genes are shaded yellow except for exons which are red and also appear as black boxes along the upper X-axis. Arrows above genes show the direction of transcription. The three CNSs located within *Fugu RAB28* are boxed.

area, which again alludes to the significant genome expansion in the mammalian lineage that is presumed to have occurred since our divergence. It is highly unlikely that three such regions were maintained by chance. Indeed in a recent survey of evolutionarily conserved footprints at the human *DACH* locus, a number of CNSs shared by distant vertebrates were identified (Nobrega *et al.*, 2003). Of nine randomly sampled elements either flanking the *DACH* gene or located in an intron, seven were found capable of directing reporter gene expression to tissues in which *DACH* is endogenously expressed. One would anticipate therefore that the three CNSs identified within the fourth intron of *RAB28* (Figure 3.8) will transpire to be *cis*-acting regulators and, considering the maintenance of synteny in this *Bapx1* region, quite possibly exert an influence on a local gene. This gene may prove to be *RAB28* itself and in this regard, their location within an intron is sufficient to ensure a continued close association.

3.2.2.3.3 Is there any evidence that ProxB and DistB are expressed genes?

Before proceeding with investigations into ProxB and DistB as putative regulatory elements, it was necessary to gather evidence in favour of their being conserved *non-coding* sequences. The highly conserved 180bp portion of ProxB was unlikely to represent an expressed sequence, however the VISTA alignment pattern of the more extensive DistB resembled that of a typical gene composed of a few well conserved exons (see Figure 3.5). A number of approaches were adopted: NIX analysis was performed to search for potential ORFs and other features associated with coding sequences, and BLAST programs available via both NCBI and HGMP web servers were employed to compare ProxB and DistB to all available EST and cDNA databases representing multiple organisms. The results failed to produce any evidence indicating that ProxB or DistB are either genes, or expressed sequences such as transcribed mRNAs. Furthermore, the human region encompassing ProxB and DistB, which formed the base sequence in all VISTA and Pip alignments, was masked for repeats using RepeatMasker (available via http://ftp.genome.washington.edu/; Smit, A. and Green, E.; unpublished) to preclude

the possibility of repeats accounting for high levels of conservation. In conclusion, ProxB and DistB appear to be genuine conserved *non-coding* sequences as no evidence was found to suggest that they are expressed.

3.2.3 Conclusions and Discussion

Previous studies have demonstrated conservation of Bapx1 at the amino acid level between human, mouse, *Xenopus* and *Drosophila* (Tribioli *et al.*, 1997). The present analysis is extended to include chicken, *Fugu* and zebrafish and serves to emphasise the high degree of conservation within a number of protein domains. There also exist additional regions of conservation of which one appears to be vertebrate-specific. It is possible to speculate that this vertebrate-specific domain may be partially responsible for evolutionary differences in functions of the invertebrate versus vertebrate *Bapx1* genes. Amino acids conserved between all species, some of which lie outside defined domains, are plausibly involved in the folding of the peptide and/or an evolutionarily conserved *in vivo* function, perhaps mediating the binding to a common interactor.

A phylogenetic tree produced to examine the evolutionary relatedness of NK3 proteins demonstrates that the Nkx3.2 sequences form a cluster within the tree separate from the amphibian Nkx3.3 orthologues and the cluster of Nkx3.1 proteins. Each cluster of orthologous sequences supports the accepted phylogeny of the species from which they derive. The invertebrate *Drosophila* bap protein lies on a separate branch to the vertebrate NK3 paralogues, as would be expected if it constitutes the ancestral NK3 peptide sequence. The evolutionary conservation of Bapx1 is indicative of its common importance during development and suggests that it may play a similar role in shared tissues of diverged species. This in turn increases the likelihood of there being conserved regulatory elements that may be detected by cross-species comparisons.

The ability to effect cross-species comparative sequence analysis is limited by the availability of sufficient high quality genomic sequence for the species concerned. Extensive *Bapx1*-containing genomic sequence was accessible for

human, mouse, *Fugu* and zebrafish which importantly traverses 450 myr of evolution. Genomic synteny in the *Bapx1* region spanning *NM_148894* (upstream) to *RAB28* (downstream) is conserved between human, mice and fish. This preservation of gene order is suggestive of functional constraint operating and raises the possibility of there being *Bapx1* regulatory elements contained within this region in a particular position relative to the gene. As expected from the sequencing of the *Fugu* genome (Aparicio *et al.*, 2002), the *Bapx1* region in fish is considerably more compact than its mammalian counterpart.

A pairwise alignment between human and mouse identified two putative *Bapx1* regulatory elements referred to as ProxB and DistB positioned 11kb and 18kb downstream of *Bapx1* respectively. A 180bp portion of ProxB was subsequently found to be conserved in zebrafish, sharing 68.33% nucleotide identity. It is somewhat surprising that none of the sequence alignment programs used were able to detect ProxB in *Fugu* considering the much closer relationship of zebrafish to *Fugu* than to the phylogenetically distant mammalian lineage. Nevertheless, *Fugu* diverged from zebrafish approximately 140 myr ago (Hedges and Kumar, 2002) during which time numerous nucleotide changes could have accumulated within the ProxB equivalent region. The identification of functional elements in non-coding DNA sequences is complicated by the fact that these elements are typically short (6-15bp) (Carroll *et al.*, 2001). Thus the BLASTZ alignment algorithm may not be sufficiently sensitive to detect a ProxB CNS in *Fugu* if only the core element was maintained. Factoring in the possibility of sequencing errors in this region could account for the inability to identify ProxB in *Fugu*.

Detection of conserved sequences between different species is profoundly influenced by a number of different factors including the chosen alignment software, length of target sequence and evolutionary distance separating the species. The latter is determined both by phylogenetic distance (years since divergence) and by evolutionary rates within a lineage or at a particular locus. MultiPipMaker was found to be more sensitive than VISTA in multi-species alignments and indeed only the former detects ProxB in zebrafish. VISTA is too stringent to identify some of the exons conserved to teleosts in mammal-fish comparisons that PipMaker is able to detect. Indeed, although the second, homeobox-containing *Bapx1* exon is clearly

present, VISTA only barely detects exon1 in *Fugu* when the *NM_148894-RAB28* region is compared (data not shown). It is perhaps, therefore, not surprising that conservation of ProxB in fish is not detected by VISTA. Interestingly, PipMaker also identified two conserved regions in *Fugu* corresponding to exons16 and 17 of *Q9P2L9*, indicating that this gene may in fact be present in *Fugu* and further corroborating conservation of synteny (see Figure 3.8). Limitations of Ensembl BLAST analysis, which failed to detect *Q9P2L9*, are thus highlighted. On the other hand, it is Ensembl BLAST alignment and not the equivalent NCBI program that detects the significant similarity between ProxB in mammals and zebrafish. The aforementioned variation between different software programs serves to illustrate the caution that one must adopt when interpreting results and the necessity of embracing a multifarious approach.

DistB does not appear to be conserved in either *Fugu* or zebrafish. However, this may simply reflect a limitation of *in silico* comparative sequence analysis. DistB was considered as a whole 8kb element in alignments and BLAST analyses, whereas it may equally be a cluster of shorter, independent regulatory elements. Indeed, as mentioned previously, elements are typically short (Carroll *et al.*, 2001) and Figure 3.6 supports the existence of *several* independent, very highly conserved regions in DistB between human and mouse, comparable to that seen for ProxB. Several of these are promising candidate regulatory regions but predicting their existence in fish is likely hampered by the difficulty of distinguishing a *bona fide* 'signal' (conservation) from background 'noise' (for example chance identity of a 6-15bp region). Alternatively the absence of DistB or parts thereof in *Fugu* and zebrafish may be genuine. It is plausible that *Bapx1* regulatory elements exist that relate specifically to a particular aspect of mammalian development. The apparent acquisition of additional roles during evolution was discussed earlier (see Introduction). Sequence conservation beyond human/mouse would not be expected in this instance.

Interestingly, a recent examination of the *HoxD* cluster has revealed the existence of a global control region occupying approximately 40kb in mammals and containing a regulatory activity that exerts an influence on multiple genes (Spitz *et al.*, 2003). Two blocks of conservation, occupying between 4kb and 5kb each, exist

at either end of this DNA segment, for which the human/mouse sequence identity is particularly high and conservation to *Fugu* is evident. The VISTA alignment of these sequence blocks resembles the human/mouse DistB VISTA alignment profile and thus provides further support for a putative regulatory potential harboured by DistB.

Importantly, no evidence was found to suggest that either DistB or ProxB are expressed sequences. Together with conservation of ProxB to zebrafish, and the extent and nature of human/mouse DistB conservation, there existed a sound foundation on which to pursue investigation of ProxB and DistB as candidate *Bapx1* regulatory elements. The next logical step was therefore to progress from speculative *in silico* analyses to a transgenic bioassay to unambiguously determine whether ProxB or DistB harboured regulatory elements.

Chapter 4 Analysis of transgenic mice

4 Analysis of transgenic mice

4.1 Introduction

Chapter 3 details the identification of two putative *Bapx1* regulatory elements, DistB and ProxB, which were initially identified on the basis of prominent shared homology between human and mouse. DistB encompasses 8kb within which there exist several shorter regions which are extremely well conserved between human and mouse. ProxB covers approximately 2kb and contains a 180bp segment that is conserved in zebrafish. The preservation of a CNS between phylogenetically distant species is a good indication that the conserved sequence is functional and subject to selective constraint. Indeed this concept has been successful in the identification of numerous regulatory elements in the past (Aparicio *et al.*, 1995; Santini *et al.*, 2003).

Following their identification, the next step was to determine whether DistB and ProxB were capable of driving gene expression in a *Bapx1*-like manner. Candidate regulatory elements are traditionally assessed via their ability to drive expression of a reporter gene when cloned into a minimal promoter reporter construct. These constructs consist of a gene whose protein product is readily assayed, such as *LacZ* (β galactosidase), linked to a minimal promoter which, by itself, is insufficient to drive reporter gene expression. The candidate sequence is cloned into the construct and the modified construct used to make transgenic animals or cell lines. A regulator is revealed by its ability to drive expression of the reporter gene, via the minimal promoter, in an appropriate pattern.

The *Bapx1* expression pattern has been extensively studied in mice, which are also the most relevant model organism to humans in terms of gut development. Accordingly, the aim was to introduce a *LacZ* minimal promoter reporter construct containing either mouse-derived DistB or ProxB into mice. Evaluation of DistB and ProxB activity would be accomplished by the harvesting and subsequent Xgal staining (to detect *LacZ* expression) of transgenic embryos at various developmental stages.

4.2 Results

4.2.1 The cloning of ProxB and DistB

The draft version of the mouse sequence was such that there existed various sequence gaps scattered throughout the *Bapx1* flanking genomic regions. The simplest and quickest method of cloning the putative regulators was thus envisaged to be by PCR using 'hybrid' primers which contained a restriction site compatible with the reporter construct at their termini. This approach was not expected to prove problematic for the 2kb ProxB sequence, however DistB, covering 8kb, was not so straightforward. A number of attempts to amplify this region using the Expand Long Template PCR System (Roche) failed, therefore standard cloning by restriction enzyme digestion was deemed to be a more viable option.

4.2.1.1 Attempts to isolate the entire 8kb DistB CNS

With the help of Dr. Martin Taylor, *in silico* analysis of the Mouse Genome Sequencing Consortium (MGSC) 3 assembly (downloaded from Ensembl) identified two BAC clones containing DistB but not *Bapx1*. BAC202D7 and BAC465K17 from the RPCI23 library were subsequently obtained from HGMP. For *Bapx1* and DistB to be separated, DistB must lie at the end of the BAC genomic insert, the idea being that this may render it easier to find a suitable restriction enzyme for DistB excision. As a result of the BAC ends having been sequenced, precise determination of the genomic insert contained in each BAC was possible and the exact location of DistB was mapped. *Bgl*II was a conveniently situated restriction enzyme for which no cleavage site was found within the 8kb; sequence analysis predicted excision of DistB within a 9348bp band. This could subsequently be subcloned into *Bam*HI-cleaved λ ZAP express (Stratagene) since *Bgl*II and *Bam*HI cleave to yield compatible 3' sites. Direct cloning into the reporter construct was not possible since *Bam*HI cleaves within *LacZ*.

Southern blot hybridisation of *Bgl*II-digested BAC DNA performed independently with probes directed against three different segments of DistB revealed that, contrary to expectation, the 9348bp band contained an internal *Bgl*II site which

consequently separated DistB into 2.9kb and 6.4kb bands (data not shown). Close inspection of the sequence further corroborated the Southern blot results: a 100bp sequence gap in DistB (the only missing sequence within DistB) was located precisely such that if it contained the *Bgl*II recognition sequence, the 9348bp DistB-containing band would be separated into 6.4kb and 2.9kb bands. Attempts were made to isolate the two aforementioned DNA bands from a gel in order to clone DistB in two separate blocks, however this strategy was constantly hampered by insufficient recovery of DNA.

4.2.1.2 Cloning by PCR

Owing to the inability to clone DistB as an entire element, the 8kb was divided into eight distinct 'homology blocks' according to the existence of conserved sequence segments between human and mouse originally identified in the VISTA alignment (see Chapter 3). It was likely that DistB actually consisted of a number of shorter control elements, perhaps individually responsible for directing spatially- or temporally-restricted gene expression in a specific tissue. Given the close proximity of the putative control elements, it was anticipated that they would affect expression of the same gene. Of course each putative internal element was not limited to exerting a stimulatory effect. DistB may in fact comprise a collection of enhancer and repressor elements whose synergistic effects ultimately determine the expression pattern of their target gene. Subdivision of DistB for the purpose of cloning and assaying enhancer activity would address the potential existence of modular regulatory elements.

Accordingly, a number of primer sets corresponding to the mouse sequence were defined with the ultimate aim being to amplify representative DistB 'homology blocks' (Figure 4.1). One of these primer sets encompassed the entire 8kb however, as mentioned previously, PCR amplification was not successful in this case. A single primer set was designed to amplify the mouse 2kb ProxB CNS; it was deemed prudent to include all of the conserved mammalian sequence in the initial assessment of enhancer activity (Figure 4.1). Confirmation of regulatory function would allow for deletion analysis of the region which could subsequently be restricted to the 180bp sequence conserved to zebrafish. The *Kpn*I recognition site (GGTACC), preceded by

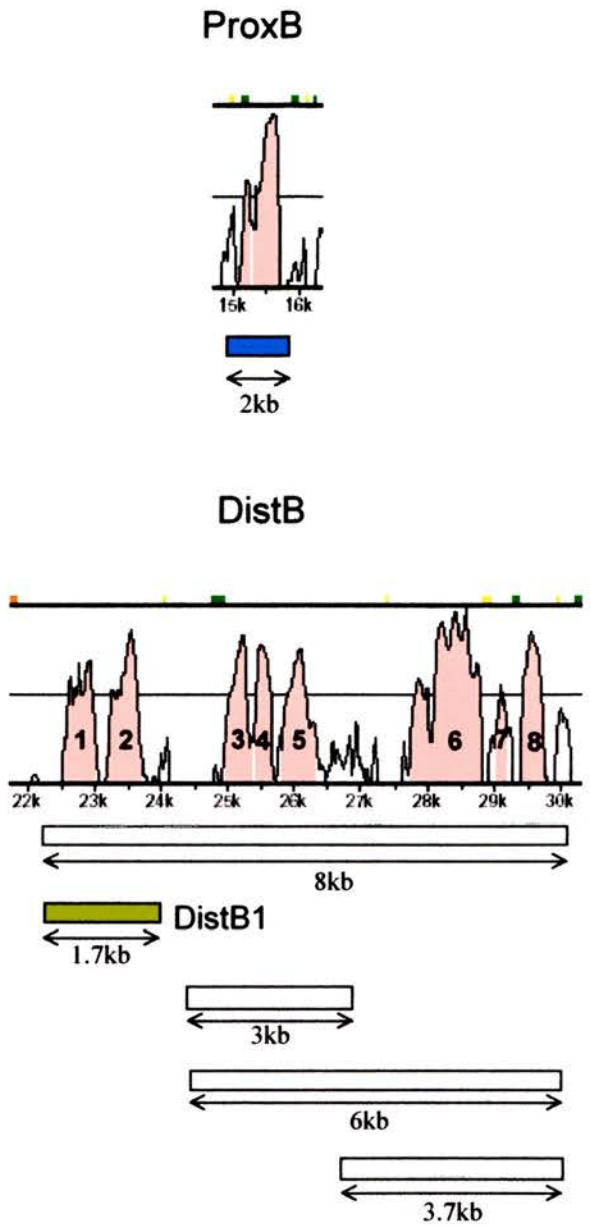


Figure 4.1: A schematic to illustrate the PCR amplification of ProxB and DistB1 from mouse genomic DNA. The ProxB and DistB CNSs appear as in Figure 3.5 and DistB homology blocks are numbered. Boxes indicate regions for which PCR amplification was attempted and the expected band sizes are shown beneath each box. Shaded boxes denote amplification reactions that were successful.

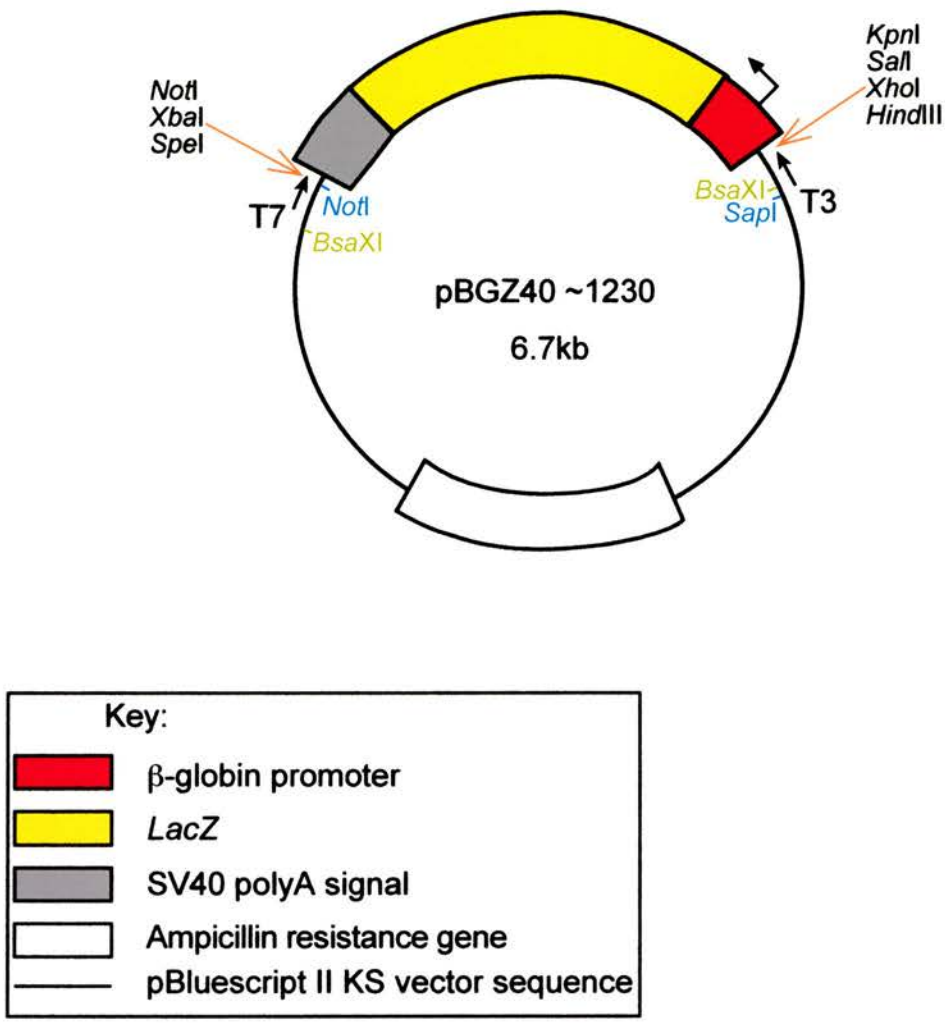


Figure 4.2: A diagram of the *LacZ* reporter construct into which ProxB and DistB1 were cloned at the *Kpn*I site. The restriction endonucleases with a single site only are indicated in the polylinker regions (orange arrows). The sites of restriction enzymes used to linearise the ProxB and DistB1 transgenes are highlighted in blue and green respectively. The approximate locations of the T3 and T7 primer sequences are indicated.

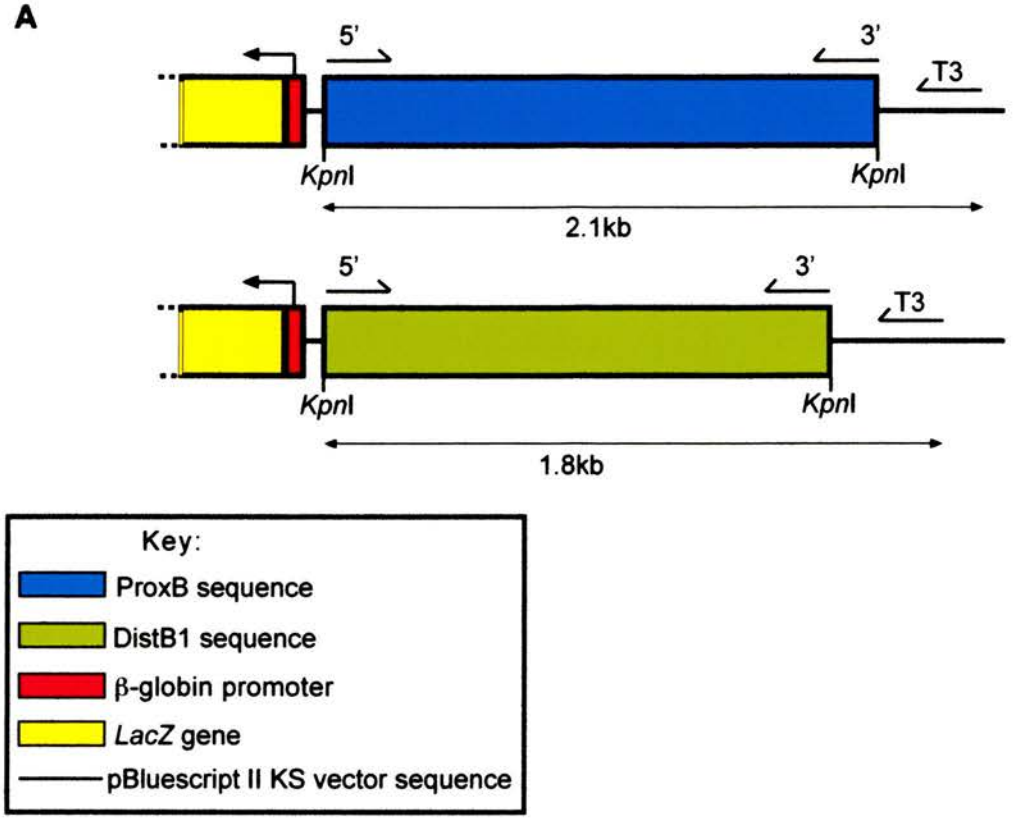
four random linker nucleotides (GATC), was incorporated into all primers at the 5' end. The *Kpn*I site is present as a single copy in the multicloning site of the *LacZ* reporter construct (Figure 4.2) and is absent from both DistB and ProxB sequence.

Despite numerous attempts and a variety of approaches, only one primer set, encompassing 'homology blocks' 1 and 2 of DistB, amplified successfully (see Figure 4.1). This portion of the DistB region will henceforth be referred to as DistB1. The GC-rich nature of sequence in the vicinity of *Bapx1* may be a contributory factor in the limited success of the PCR, which may also be influenced by the length of the target sequence; only the DistB1 primer set target was substantially less than 3kb (1.7kb). ProxB (2kb) was successfully amplified. Both ProxB and DistB1 were ligated to T-Easy vectors and transformed into either DH5α or BL21 cells. DNA was harvested from transformants identified as containing ProxB/DistB1 prior to *Asp*718 digestion to excise the 2kb/1.7kb bands respectively. *Asp*718-cleaved ProxB/DistB1 bands were ligated to *Asp*718-cleaved, purified and alkaline phosphatased *LacZ* reporter constructs. These modified reporter constructs were cloned by transformation into electrocompetent XL-1 blue cells and clones containing the reporter construct selected for by their acquisition of Ampicillin (Amp) resistance.

4.2.1.3 Verification of ProxB and DistB1 presence

This was accomplished in three ways. Firstly, a selection of Amp resistant colonies was grown up overnight, DNA harvested and reporter constructs containing ProxB/DistB1 identified by *Asp*718 excision of an appropriately sized band (data not shown). Secondly, ProxB/DistB1 inserts in T-Easy vectors were sequenced to ensure that the correct PCR product had been amplified and cloned. Approximately 500bp sequence was obtained in each case and found to be identical to the Ensembl-derived sequence. Thirdly, diagnostic PCRs were undertaken using the modified *LacZ* reporter construct as substrate in order to determine the orientation of the 1.7kb/2kb bands with respect to the direction of *LacZ* transcription (Figure 4.3A). Enhancers function in an orientation-independent manner with respect to their target gene (Lewin, 2000), nevertheless it was thought that this information may prove useful when interpreting results and the PCR reactions also served to confirm the cloning of

A



B

ProxB constructs

Construct	5' primer+T3	3' primer+T3
1	✔	
2	✔	
3		✔
4		✔
5	✔	
6		✔

DistB1 constructs

Construct	5' primer+T3	3' primer+T3
1		✔
2		✔
3		✔

Figure 4.3: A schematic to illustrate the PCR assay used to determine the orientation of the ProxB and DistB1 sequences with respect to the direction of *LacZ* transcription. (**A**) Only one of the two primers originally used to amplify the ProxB/DistB1 sequences will function in combination with the T3 primer to yield a 2.1kb/1.8kb PCR product respectively. The 5' (forward) and 3' (reverse) designation of primers relates to the orientation in which the ProxB and DistB1 sequences appear in Figure 3.5. (**B**) A summary table of PCR orientation assay results. ProxB was cloned in both orientations relative to the direction of *LacZ* transcription whilst the three DistB1 constructs all contain DistB1 in the opposite orientation with respect to *LacZ* transcription direction.

the correct sequence into the reporter construct. In interpreting the PCRs, I define ProxB and DistB1 as running 5' to 3' in the same direction as they appear in Figure 3.5 in relation to their upstream (*Bapx1*) and downstream (*RAB28*) flanking genes. In this respect, DistB1 was cloned into the reporter construct in the opposite orientation to *LacZ*, whilst ProxB was inserted in the same orientation as *LacZ* in three instances and opposite orientation in a further three cases (Figure 4.3B).

4.2.2 Preparation of the transgene

To ensure adequate amounts of clean DNA were produced for microinjection, the ProxB and DistB1 transgenes were prepared from DNA harvested from larger volume cultures. (ProxB constructs contained the element in the same orientation as *LacZ*; DistB1 elements were in the opposite orientation. See above). The linearised transgenes are depicted diagrammatically in Figure 4.4. The DistB1 construct (containing the 1.7kb band) was linearised with *Bsa*XI, and the 6kb transgene purified away from pBluescriptKS vector DNA. Due to *Bsa*XI cleaving within ProxB, a *Not*I/*Sap*I double digest was employed in this case to separate the linearised transgene from superfluous vector DNA. A diagnostic *Asp*718 digestion confirmed that in both cases the correct bands were cloned into the reporter construct (data not shown). The linearised ProxB and DistB1 transgenes were microinjected into recipient F1 fertilised oocytes at a concentration of 2ng/μl. The microinjections, subsequent oviductal transfers and initiation of transgenic lines were carried out by Lorna Purdie. Eight offspring from ProxB microinjections were positively identified as containing the transgene by a *LacZ* genotyping PCR whilst the number of positive offspring for DistB1 was six (data not shown). All 14 offspring constituted founders of transgenic lines. Table 4.1 provides a summary of the steps leading to the initiation of the ProxB and DistB1 transgenic lines. Embryos derived from ProxB and DistB1 lines were harvested at various developmental stages and subjected to Xgal staining to assess the activity of the ProxB 2kb and DistB1 1.7kb CNSs. The data presented below indicates that both ProxB and DistB1 are regulatory elements, capable of directing tissue-specific gene expression during mouse development. Furthermore, detailed analysis of the resultant expression patterns suggests that they may contribute to the endogenous regulation of *Bapx1*.

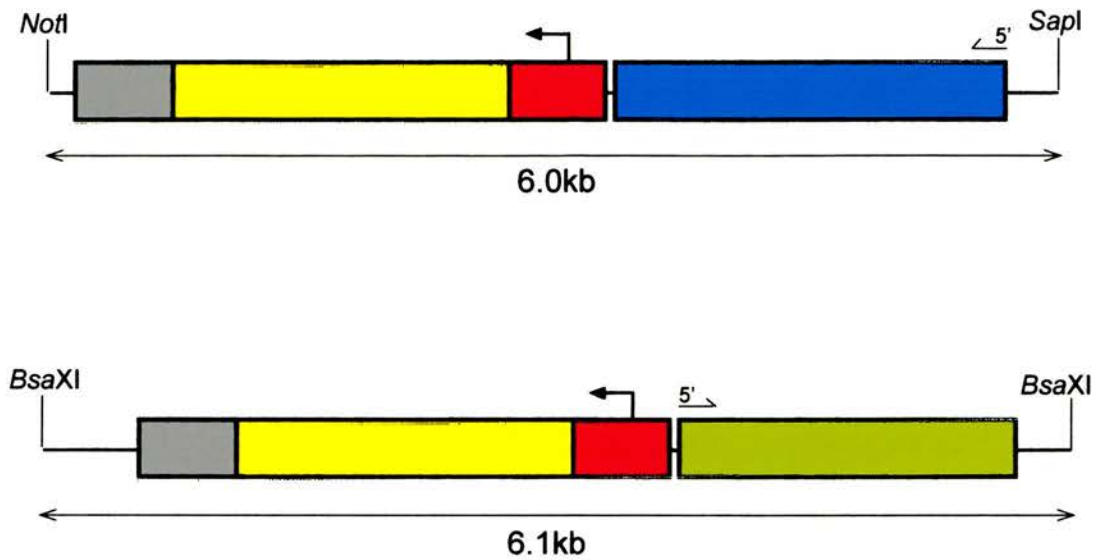


Figure 4.4: A diagram to illustrate the linear ProxB and DistB1 transgenes. The restriction enzymes used to excise each transgene from the reporter construct are indicated at each end of the transgene. The orientation of the ProxB and DistB1 sequences is shown by the position of the 5' (forward) primer; direction was originally designated according to the appearance of the ProxB and DistB1 sequence in Figure 3.5.

	ProxB	DistB1
Total number of fertilised oocytes	NR	376
Total number of fertilised oocytes injected with the transgene	NR	260
Total number of cells cultured	NR	224
Total number of 2 cell-stage embryos transferred to oviducts	NR	193
Number of embryos born	45; 1 died	30
Number of transgenic embryos amongst those born	8	6
Number of transgenic F1 lines transmitting	5	6
Number of lines showing reporter gene expression	3	4

Table 4.1: A summary table detailing steps leading to the initiation of the ProxB and DistB1 transgenic lines. NR, not recorded.

4.2.3 Expression pattern directed by ProxB

Eight mice were positively identified as carrying the ProxB CNS and thus were intended to constitute the founders of transgenic lines. Of these, one female failed to breed and one male was infertile, whilst a further line did not transmit the transgene. Embryos from the remaining five (first generation) transgenic lines (termed A267, A269.1, A269.2, A269.3 and A269.5) were harvested between E8.5 and E16.5 days of gestation. Lines A269.1 and A269.3 did not show any signs of reporter gene expression at any of the stages examined. The other three lines all exhibited identical staining patterns throughout development and are therefore considered collectively. ProxB-directed reporter gene expression was restricted to a specific domain of the upper embryo. Parallels are drawn with endogenous *Bapx1* expression. A panel to illustrate the endogenous *Bapx1* expression pattern in wholemount embryos between the ages of E9.5 and E13.5 is included in Appendix 2.

4.2.3.1 Early embryonic development

Here I include stages E8.5 through E11.5. No expression is detectable at E8.5, however by E9.5, faint staining is visible in the inferior section of the mandibular portion of the first branchial arch (Figure 4.5A). This is reminiscent of a novel domain of *Bapx1* expression that appears in E9.5 embryos in neural crest-derived mesenchyme of the first branchial arch (Tribioli *et al.*, 1997). By E10.5, expression is considerably stronger (Figure 4.5B, C) and sagittal sections demonstrate that, within the lower half of the first branchial arch, staining intensity is greatest in the most inferior region (Figure 4.5D). Transverse sections at E10.5 further reveal that expression is confined to a discrete domain of the first branchial arch (Figure 4.5E). This region of expression closely reflects that of endogenous *Bapx1* at E10.5 (Figure 4.5B'; R. Watson, unpublished data). By E11.5, the expression domain is beginning to expand within the mandibular component of the first branchial arch as it develops, though it remains in the inferior portion (Figure 4.5F, G).

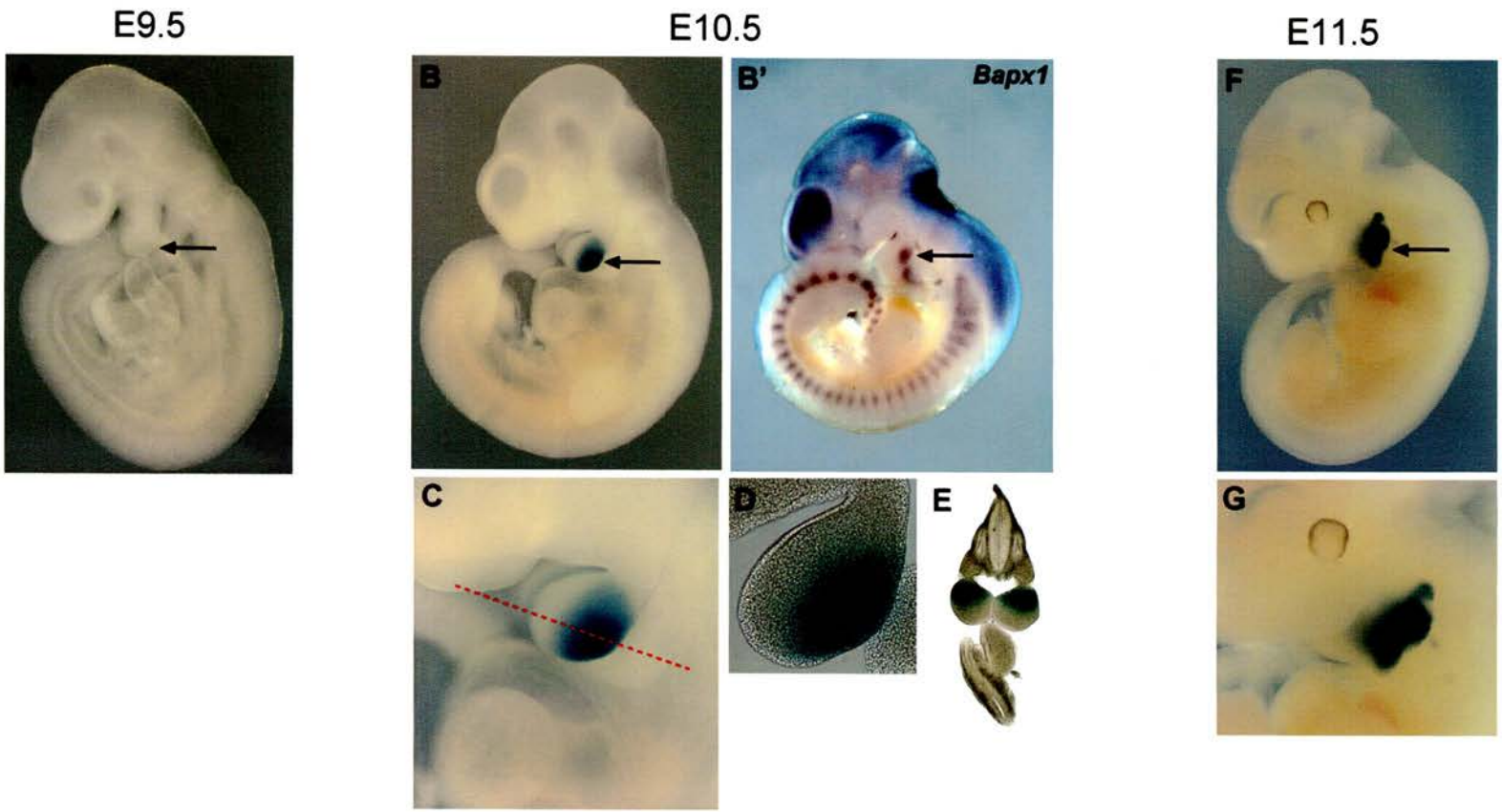


Figure 4.5: Localisation of ProxB-driven reporter gene expression by Xgal staining during early stages (E9.5-E11.5) of mouse embryogenesis. Reporter gene expression localises to the first branchial arch (arrow) at E9.5 (**A**), E10.5 (**B-E**) and E11.5 (**F**, **G**). The expression domain in the first branchial arch at E10.5 is comparable to that of endogenous *Bapx1* (**B'**; image provided by Dr. Robert Watson). Sagittal (**D**) and transverse sections (**E**) at E10.5 confirmed the expression domain to be the mandibular portion of the first branchial arch. The red dotted line in **C** indicates the section angle for **E**. By E11.5, the expression domain has begun to expand within the branchial arch region as development proceeds (**F** and close up in **G**).

4.2.3.2 Mid embryonic development

The expression pattern at E12.5 appears to be very similar to that at E11.5 (compare Figures 4.6A(i), (ii) and 4.5F, G) although by E12.5, derivative structures of the branchial arches are better defined. Development of the branchial arch region has now given rise to identifiable separate structures such as the tongue, jaw and auditory apparatus. An internal view looking up into the head reveals two distinct, symmetrical expression domains; a small staining patch which appears to correspond to the region of the future middle ear and a significantly wider and more extensive domain following the line of the jaw (Figure 4.6A(iii)). Transverse sections confirm these observations, revealing that the jaw domain covers the primordia of two of the middle ear ossicles (the malleus and incus) in addition to the precartilage primordium of Meckel's cartilage within the lower jaw (Figure 4.6D-F). Sagittal sections of E12.5 embryos demonstrate staining in the lower jaw/Meckel's cartilage region (Figure 4.6B(i), (ii)).

By E13.5, individual structures are more clearly defined and the external staining visible is less extensive but still very strong in the ear region (Figure 4.6C(i), (ii)). An internal view of a dissected head is more informative, revealing that expression following the line of the jaw is also very distinct (Figure 4.6C(iii)). The expression domain has clearly lengthened, and narrowed somewhat in the distal Meckel's cartilage regions, between E12.5 and E13.5 as jaw development proceeds. The strongest expression is seen proximal to the ear. Notably, *Bapx1* expression has been reported in both the primordia of middle ear-associated bones and Meckel's cartilage (Tucker *et al.*, 2003, submitted). The level of endogenous *Bapx1* expression in the latter is very low, however, and is not detected along the lower jaw line. However, in mice which possess a 'knock-in' of a *LacZ* reporter gene targeted to the *Bapx1* locus and replacing the entire coding region ($Bapx1^{LacZ}$) (Akazawa *et al.*, 2000), *LacZ* expression is detected along the mandible bone of the lower jaw (see later). At E13.5, ProxB-driven *LacZ* expression is also apparent in the submandibular (salivary) glands (indicated in Figure 4.6C(iii) by short protrusions emanating from the straight blue lines of stain in the jaw).

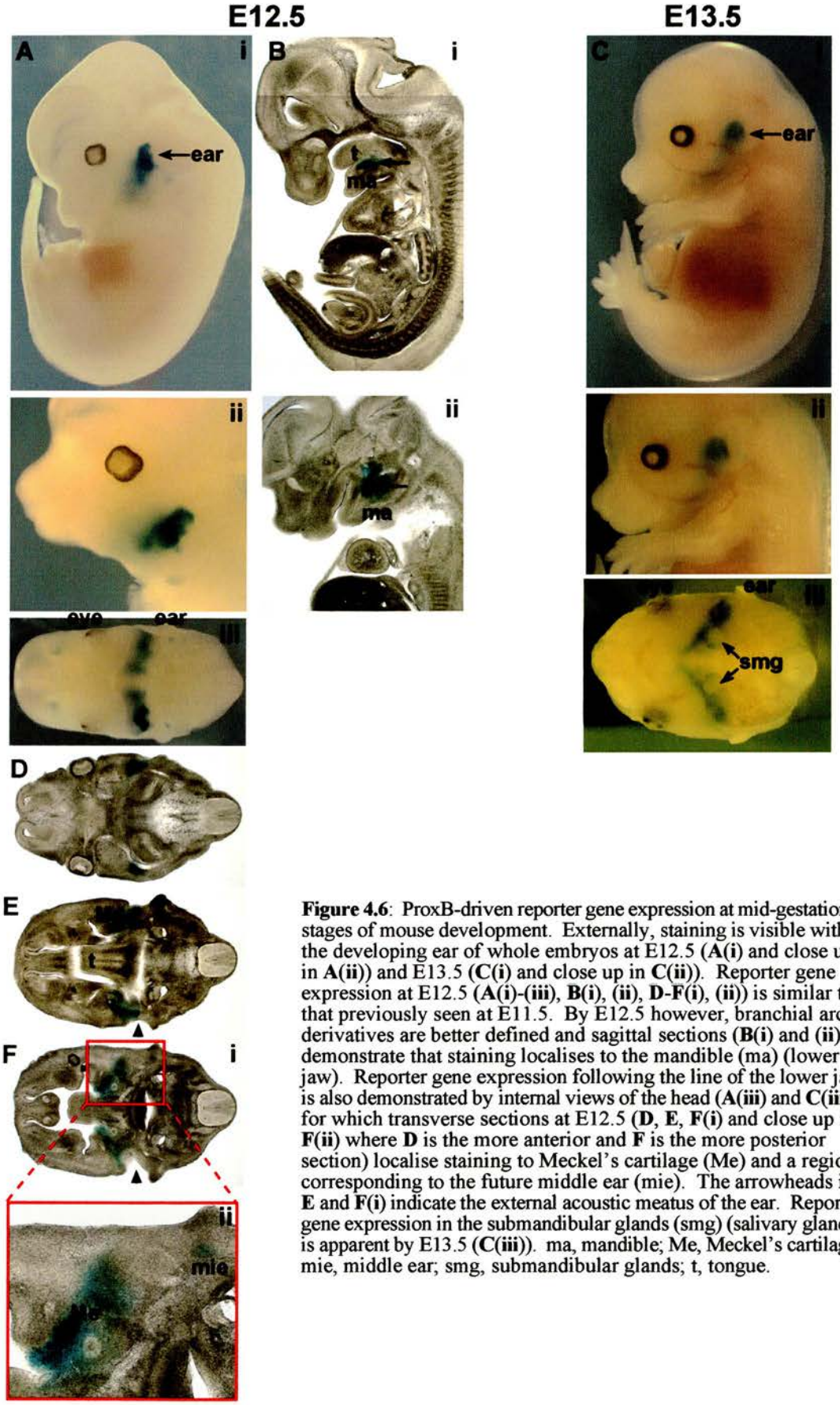


Figure 4.6: ProxB-driven reporter gene expression at mid-gestational stages of mouse development. Externally, staining is visible within the developing ear of whole embryos at E12.5 (**A(i)** and close up in **A(ii)**) and E13.5 (**C(i)** and close up in **C(ii)**). Reporter gene expression at E12.5 (**A(i)-(iii)**, **B(i)**, **(ii)**, **D-F(i)**, **(ii)**) is similar to that previously seen at E11.5. By E12.5 however, branchial arch derivatives are better defined and sagittal sections (**B(i)** and **(ii)**) demonstrate that staining localises to the mandible (ma) (lower jaw). Reporter gene expression following the line of the lower jaw is also demonstrated by internal views of the head (**A(iii)** and **C(iii)**), for which transverse sections at E12.5 (**D**, **E**, **F(i)** and close up in **F(ii)** where **D** is the more anterior and **F** is the more posterior section) localise staining to Meckel's cartilage (Me) and a region corresponding to the future middle ear (mie). The arrowheads in **E** and **F(i)** indicate the external acoustic meatus of the ear. Reporter gene expression in the submandibular glands (smg) (salivary glands) is apparent by E13.5 (**C(iii)**). ma, mandible; Me, Meckel's cartilage; mie, middle ear; smg, submandibular glands; t, tongue.

4.2.3.3 Late embryonic development

Reporter gene expression at the later developmental stages examined is extremely similar to the pattern observed at E13.5. Embryos harvested at E14.5, E15.5 and E16.5 days of gestation all exhibit highly specific *LacZ* staining in the head region only, as in previous developmental stages. Externally, a patch of expression in each ear is visible, whilst analysis of dissected heads reveals that internal staining follows the line of Meckel's cartilage along the jaw (Figures 4.7 and 4.8). Staining in the submandibular glands, which first became apparent at E13.5, continues to be detected at these later developmental stages. Development and lengthening of Meckel's cartilage as jaw development proceeds is faithfully reflected by *LacZ* staining between E14.5 and E16.5, which demonstrates that by E16.5, the two symmetrical lines of expression (right and left Meckel's cartilage) have fused at the tip (Figure 4.8F).

The most complete analysis of ProxB embryos was conducted at E14.5. Dissected heads were dehydrated in 100% MeOH overnight and 'cleared' in BABB solution (as described in Materials and Methods) which renders tissue more transparent, allowing visualisation of internal structures without further dissection (Figure 4.7D). This clearly demonstrates reporter gene expression extending from the most intense region in the ear, along Meckel's cartilage. This continuous expression domain encompasses that of the developing middle ear (in which *Bapx1* is expressed; Tucker *et al.*, 2003, submitted), with proximal ear staining as seen in Figure 4.7D corresponding closely to the region in which the middle ear apparatus is located (Figure 4.7D'; Mallo, 1997). Staining in salivary glands is also apparent. Treated heads were subjected to examination by Optical Projection Tomography (OPT) (see Materials and Methods) and two movies produced showing the location of *LacZ* staining in ProxB heads at E14.5 from both a frontal and a sagittal perspective (see Movie 4.1 and 4.2 in attached CD). Further confirmation of the associated ear-Meckel's cartilage-submandibular gland expression domains derives from both sagittal (Figure 4.7E, F) and transverse (Figure 4.7G-I) sections taken at E14.5.

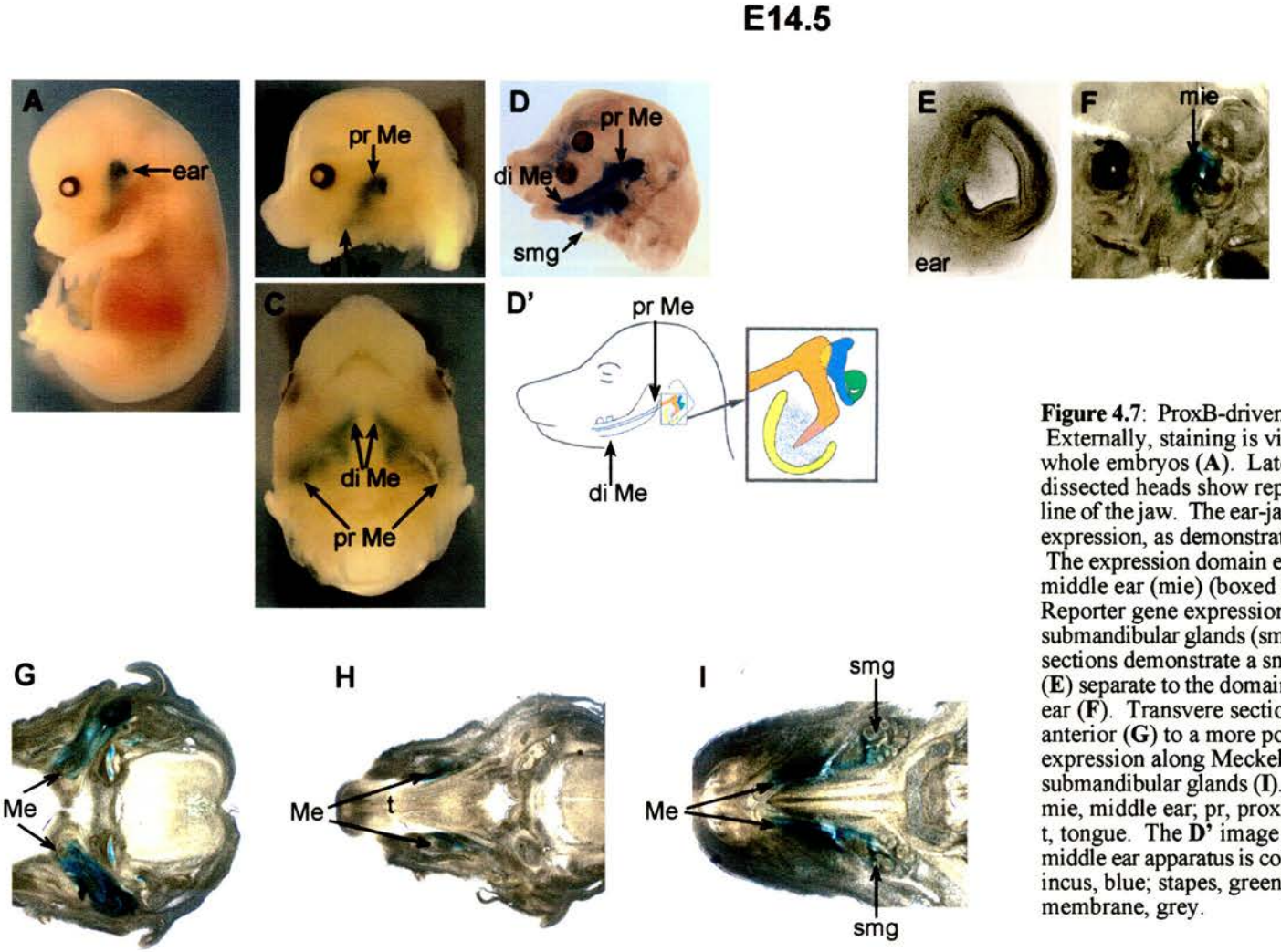


Figure 4.7: ProxB-driven reporter gene expression at E14.5. Externally, staining is visible within the developing ear of whole embryos (**A**). Lateral (**B**) and internal (**C**) views of dissected heads show reporter gene expression along the line of the jaw. The ear-jaw domain is a region of continuous expression, as demonstrated by a cleared head at E14.5 (**D**). The expression domain encompasses that of the developing middle ear (mie) (boxed in **D'**) (compare **D**, **D'** and **F**). Reporter gene expression is also detected in the submandibular glands (smg) (salivary glands) (**D, I**). Sagittal sections demonstrate a small patch of stain in the ear region (**E**) separate to the domain in the region of the future middle ear (**F**). Transvere sections through the head from a more anterior (**G**) to a more posterior (**I**) level show *LacZ* expression along Meckel's cartilage (Me) (**G-I**) and in the submandibular glands (**I**). di, distal; Me, Meckel's cartilage; mie, middle ear; pr, proximal; smg, submandibular glands; t, tongue. The **D'** image was taken from Mallo, 1997; middle ear apparatus is coloured as follows: malleus, orange; incus, blue; stapes, green; tympanic ring, yellow; tympanic membrane, grey.

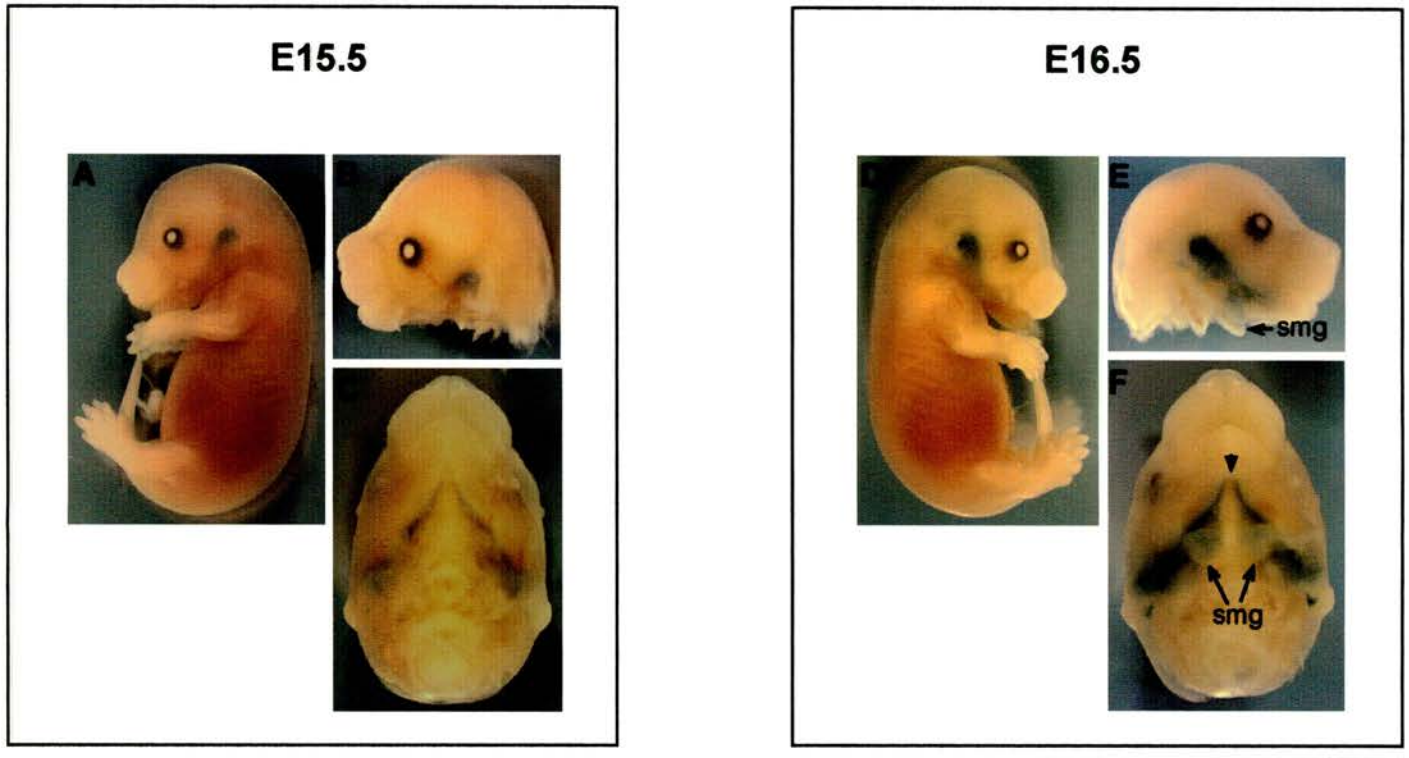


Figure 4.8: Localisation of ProxB-driven reporter gene expression by Xgal staining during late stages (E15.5 and E16.5) of mouse embryogenesis. Externally, *LacZ* expression at E15.5 (**A**, **B**) and E16.5 (**D**, **E**) reflects the staining pattern seen at E13.5 (Figure 4.6 C(i), (ii)) and E14.5 (Figure 4.7 A, B). Internal views of dissected heads demonstrate that the staining pattern faithfully reflects the lengthening of Meckel's cartilage (**C**, **F**). By E16.5, the right and left Meckel's cartilages have fused at the tip (arrowhead in **F**). Expression persists in the submandibular glands (smg) (**E**, **F**).

4.2.4 Expression pattern directed by DistB1

Transgenic DistB1 embryos were harvested from six positive lines (termed A266, A266.1, A266.2, A266.3, A266.4 and A266.5) between E8.5 and E15.5 days of gestation. A266.1 and A266.5 did not display any signs of *LacZ* reporter expression at any of the stages examined. The other four lines all demonstrated comparable reporter gene expression. The description below concentrates on line A266.3 due to its exhibiting the strongest staining pattern at all stages examined. The remaining three lines all stained in a subset of the A266.3 expressing regions and due reference is made to these conserved regions of expression. Comparisons are also drawn with endogenous *Bapx1* expression and with reporter gene expression detected in $Bapx1^{LacZ}$ mice. A panel to illustrate the endogenous *Bapx1* expression pattern in wholemount embryos between the ages of E9.5 and E13.5 is included in Appendix 2. In stark contrast to the restricted expression domain within the head driven by ProxB, DistB1-directed *LacZ* expression is relatively widespread and detected in multiple tissues of the developing mouse embryo.

4.2.4.1 Early embryonic development

In E8.5 embryos expression in the heart is clearly detectable, along with stain in the cephalic mesenchyme and a faint patch either side of the embryo positioned slightly posterior and dorsal to the developing heart (Figure 4.9A). Expression in cephalic mesenchyme is also detected in $Bapx1^{LacZ}$ mice, however DistB1 does not drive expression in the somites or splanchnic mesoderm as seen for $Bapx1^{LacZ}$ (Figure 4.9A', B'). By E9.5, expression surrounding the otic vesicle (which develops into the internal ear) and in distinct regions of the brain is apparent; transverse sections demonstrate expression in the diencephalon (midbrain) and hindbrain but not the forebrain (Figure 4.9B, C). The distinct boundaries illustrated are indicative of specific, directed expression of the transgene rather than background expression. Expression is also noted in the lateral portion of the embryo immediately anterior to the forelimb bud and faintly along the anterior tip of the forelimb bud (Figure 4.9D). Internally, expression is detected in the developing lung buds (Figure 4.9E).

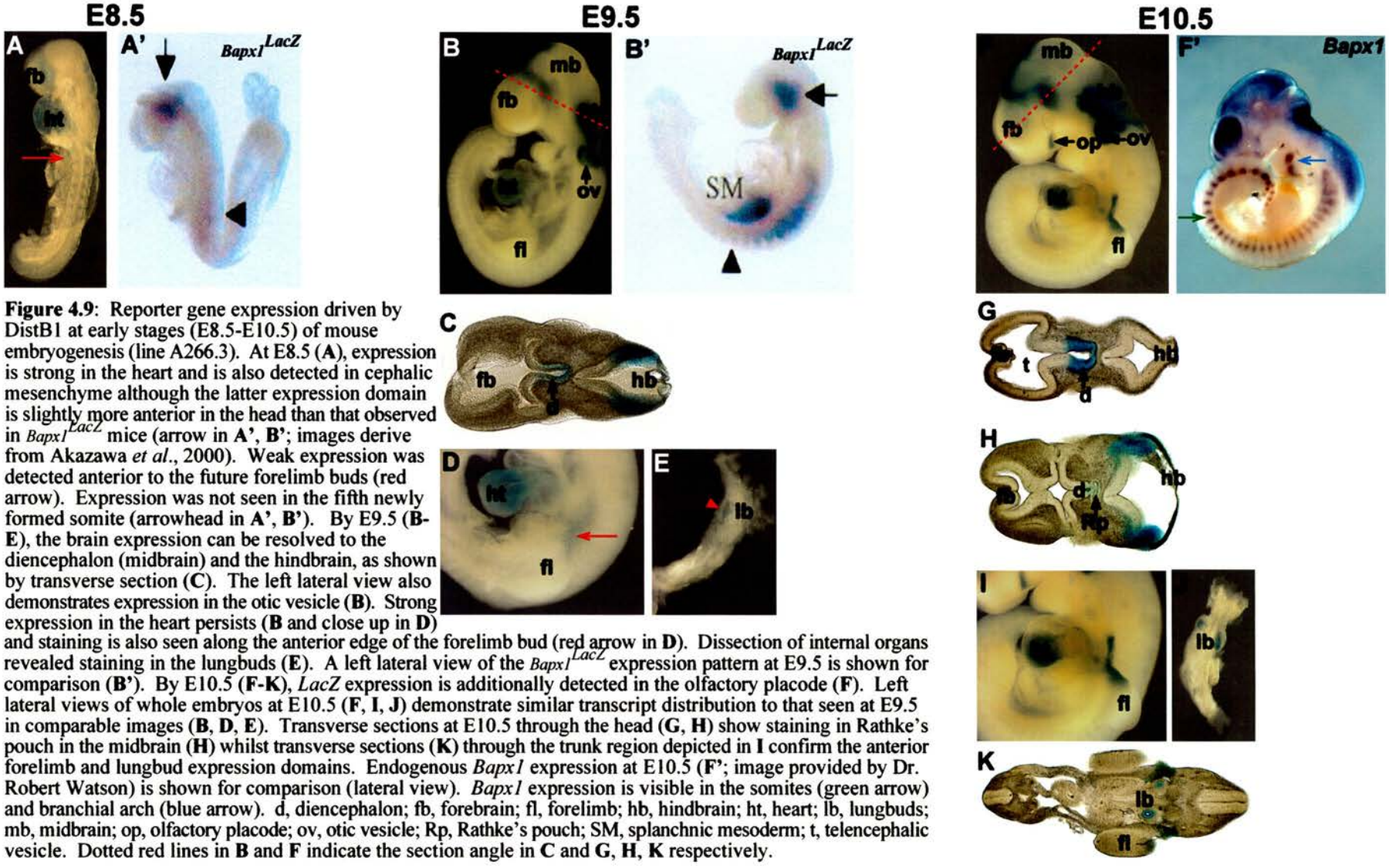


Figure 4.9: Reporter gene expression driven by DistB1 at early stages (E8.5-E10.5) of mouse embryogenesis (line A266.3). At E8.5 (**A**), expression is strong in the heart and is also detected in cephalic mesenchyme although the latter expression domain is slightly more anterior in the head than that observed in *Bapx1*^{LacZ} mice (arrow in **A'**, **B'**; images derive from Akazawa *et al.*, 2000). Weak expression was detected anterior to the future forelimb buds (red arrow). Expression was not seen in the fifth newly formed somite (arrowhead in **A'**, **B'**). By E9.5 (**B-E**), the brain expression can be resolved to the diencephalon (midbrain) and the hindbrain, as shown by transverse section (**C**). The left lateral view also demonstrates expression in the otic vesicle (**B**). Strong expression in the heart persists (**B** and close up in **D**) and staining is also seen along the anterior edge of the forelimb bud (red arrow in **D**). Dissection of internal organs revealed staining in the lungbuds (**E**). A left lateral view of the *Bapx1*^{LacZ} expression pattern at E9.5 is shown for comparison (**B'**). By E10.5 (**F-K**), *LacZ* expression is additionally detected in the olfactory placode (**F**). Left lateral views of whole embryos at E10.5 (**F**, **I**, **J**) demonstrate similar transcript distribution to that seen at E9.5 in comparable images (**B**, **D**, **E**). Transverse sections at E10.5 through the head (**G**, **H**) show staining in Rathke's pouch in the midbrain (**H**) whilst transverse sections (**K**) through the trunk region depicted in **I** confirm the anterior forelimb and lungbud expression domains. Endogenous *Bapx1* expression at E10.5 (**F'**; image provided by Dr. Robert Watson) is shown for comparison (lateral view). *Bapx1* expression is visible in the somites (green arrow) and branchial arch (blue arrow). d, diencephalon; fb, forebrain; fl, forelimb; hb, hindbrain; ht, heart; lb, lungbuds; mb, midbrain; op, olfactory placode; ov, otic vesicle; Rp, Rathke's pouch; SM, splanchnic mesoderm; t, telencephalic vesicle. Dotted red lines in **B** and **F** indicate the section angle in **C** and **G**, **H**, **K** respectively.

Staining in the olfactory placode becomes visible by E10.5 (Figure 4.9F) but otherwise, transcript distribution is similar to that seen at E9.5 (Figure 4.9I, J). Transverse sections at the level of the head again show expression in the diencephalon (Figure 4.9G). Additionally, slightly posteriorly, staining is visible in Rathke's pouch and the infundibular recess of the diencephalon as well as the hindbrain (Figure 4.9H). Transverse sections at a more posterior level of the embryo confirm expression along the anterior edge of the forelimb bud and in the lungs (Figure 4.9K). Wholemount *in situ* hybridisation with a *Bapx1* probe at E10.5 demonstrates *Bapx1* expression in the first branchial arch and somites, for which expression is stronger in more rostral (developmentally older) somites (Figure 4.9F'; R. Watson, unpublished data). No *LacZ* expression is detected in the other three expressing DistB1 transgenic lines during these early stages of development.

4.2.4.2 Mid embryonic development

Reporter gene expression at E11.5, as at E10.5, is detected in the brain, nasal processes, otic vesicle, heart, lungs and anterior edge of the forelimbs (Figure 4.10A, D). Additionally, faint staining is now also detectable in the developing stomach (Figure 4.10D). Stronger expression in the posterior stomach is evident in two of the other DistB1 lines, in one case covering almost the entire posterior stomach (Figure 4.10E) and in the other case restricted to its lower end, anterior to the pyloric sphincter (Figure 4.10F). Rotation of the latter 90° on its perpendicular axis demonstrates that expression encircles the lower posterior stomach as illustrated by the ring of blue stain (Figure 4.10G). Wholemount *in situ* hybridisation demonstrates that *Bapx1* expression in the gut is restricted to the posterior stomach and throughout the spleen (Figure 4.10C). In whole embryos at E11.5, *Bapx1* transcripts are externally visible in the branchial arch, rostral somites and as a stripe of expression marking the future digits of the fore- and hindlimbs (Figure 4.10B; R. Watson, unpublished data). A similar pattern of *Bapx1* expression is observed at E12.5 (Figure 4.10I; R. Watson, unpublished data, 4.10O).

At E12.5, *LacZ* staining is additionally detected in the dorsal root ganglia along the spine of the embryo (Figure 4.10H, Q, R). The reporter gene is also highly expressed in the developing bones of the forelimb and hindlimb. Detached forelimbs

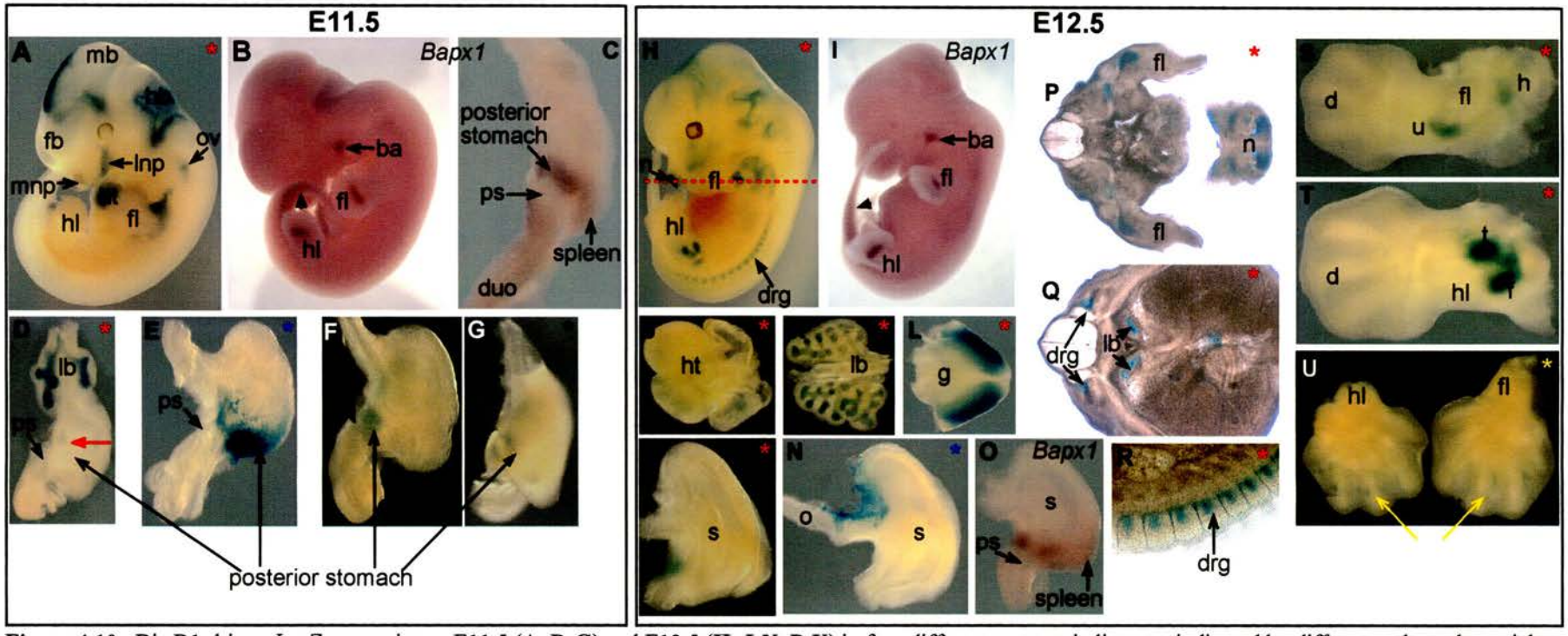


Figure 4.10: DistB1-driven *LacZ* expression at E11.5 (**A**, **D-G**) and E12.5 (**H**, **J-N**, **P-U**) in four different transgenic lines, as indicated by different coloured asterisks. The left lateral view of a whole embryo at E11.5 (**A**) demonstrates reporter gene exression in the brain, nasal processes, otic vesicle, heart and anterior edge of the forelimb. Endogenous *Bapx1* expression as seen externally in the future digits of the fore- and hindlimb, branchial arch and caudal somites (arrowheads) is shown for comparison at E11.5 (**B**) and at E12.5 (**I**). Dissection of internal organs from E11.5 embryos revealed strong reporter gene expression in the developing lungbuds (**D**) and in various domains of the posterior stomach within the gut, the pattern varying slightly between different transgenic lines (**D-G**). The red arrow in **D** points to faint expression in the posterior stomach of line A266.3, which is stronger by E12.5 (**M**). Expression encircling the lower posterior stomach of line A266 (**G**) is demonstrated by rotation of the gut in **F** 90° along its perpendicular axis. *Bapx1* expression within the foregut is restricted to the posterior stomach and spleen (**C**, **O**). (**H**) Left lateral view of an E12.5 embryo from line A266.3. At this stage, *LacZ* expression is additionally detected in the dorsal root ganglia, shown also by transverse (**Q**) and sagittal (**R**) sections. Expression is seen in mesenchymal condensations of the fore- and hindlimbs both in whole embryos (**H** and close up in **S**, **T**) and transverse sections (**P**) which appear to correspond to the humerus and ulna of the forelimb (**S**) and the tibia and fibula of the hindlimb (**T**). Expression is visible flanking the developing digits of the fore- and hindlimbs in line A266.2 (**U**) (yellow arrows). Internally, reporter gene expression persists in the heart (**J**), lungbuds (**K**, **Q**) and gut (**M**, **N**) and is also strong in the genitals (**L**). Transcripts that were detected in the posterior stomach of line A266.4 at E11.5 (**E**) are seen along the inner edge of the anterior stomach, proximal to the oesophagus at E12.5 (**N**). The angle of transverse sections (**P**, **Q**) is denoted by the red dotted line in **H**. ba, branchial arch; d, digits; drg, dorsal root ganglia; duo, duodenum; f, fibula; fb, forebrain; fl, forelimb; g, genitals; h, humerus; hb, hindbrain; hl, hindlimb; ht, heart; lb, lungbuds; lnp, lateral nasal process; mb, midbrain; mnp, medial nasal process; n, nose; o, oesophagus; ov, otic vesicle; ps, pyloric sphincter; s, stomach; t, tibia; u, ulna. * Line A266.3; * Line A266.4; * Line A266; * Line A266.2. Images **B** and **I** were kindly provided by Dr. Robert Watson.

demonstrate staining of the (proximal) humeral and (distal) ulnar mesenchymal condensations (Figure 4.10S) whilst condensations of the (distal) tibia and fibula of the hindlimb appear as two blue patches in Figure 4.10T. The extra patch of stain seen in Figure 4.10H at the proximal end of the hindlimb may reflect the chondrification centre of the femur. No staining is detected in the digits of line A266.3, however pairs of blue spots flanking the proximal phalange bones of the forelimb and hindlimb digits are visible in line A266.2 (Figure 4.10U).

Internally, in addition to maintenance in the heart, lungs and posterior stomach of the gut (Figure 4.10J, K, M), staining is clearly visible in the genitals (Figure 4.10L). In contrast to its posterior stomach expression at E11.5, staining in line A266.4 at E12.5 is confined to loose tissue along the inner curve of the anterior stomach and at the proximal end of the oesophagus (Figure 4.10N). Transverse sections of A266.3 embryos at E12.5 confirm expression in the regions previously described and two representative sections are shown in Figure 4.10P, Q.

4.2.4.3 Late embryonic development

I examined reporter gene expression during later stages of gestation in E13.5, E14.5 and E15.5 embryos. Two novel domains of expression supplementary to those seen at E12.5 become apparent at E13.5; the eyelid, and two discrete, symmetrical pairs of spots which correspond to the position of mammary gland primordia (Figure 4.11A). These novel domains of expression persist at later stages of embryogenesis (Figure 4.12A, C; 4.13A).

Analysis of whole embryos demonstrates that expression in the ear is visible for lines A266.3 and A266.2 from E13.5 (Figure 4.11A, B; 4.12C, D, G, G'; 4.13A). Another line, A266, shows staining of the dorsal root ganglia as in line A266.3, which continues to be detectable in both lines at E15.5 (Figure 4.11A, C; 4.12B, C, H, H', J; 4.13B, C). Transverse sections at E14.5 confirm expression in the dorsal root ganglia, lateral to the neural tube (Figure 4.12H, H'). This does not correspond to endogenous *Bapx1* expression which, at E13.5, is detected externally by *in situ* hybridisation in rostral somites of the vertebral column (Figure 4.11D; R. Watson, unpublished data) and internally is expressed in mesenchyme surrounding the notochord (Akazawa *et*

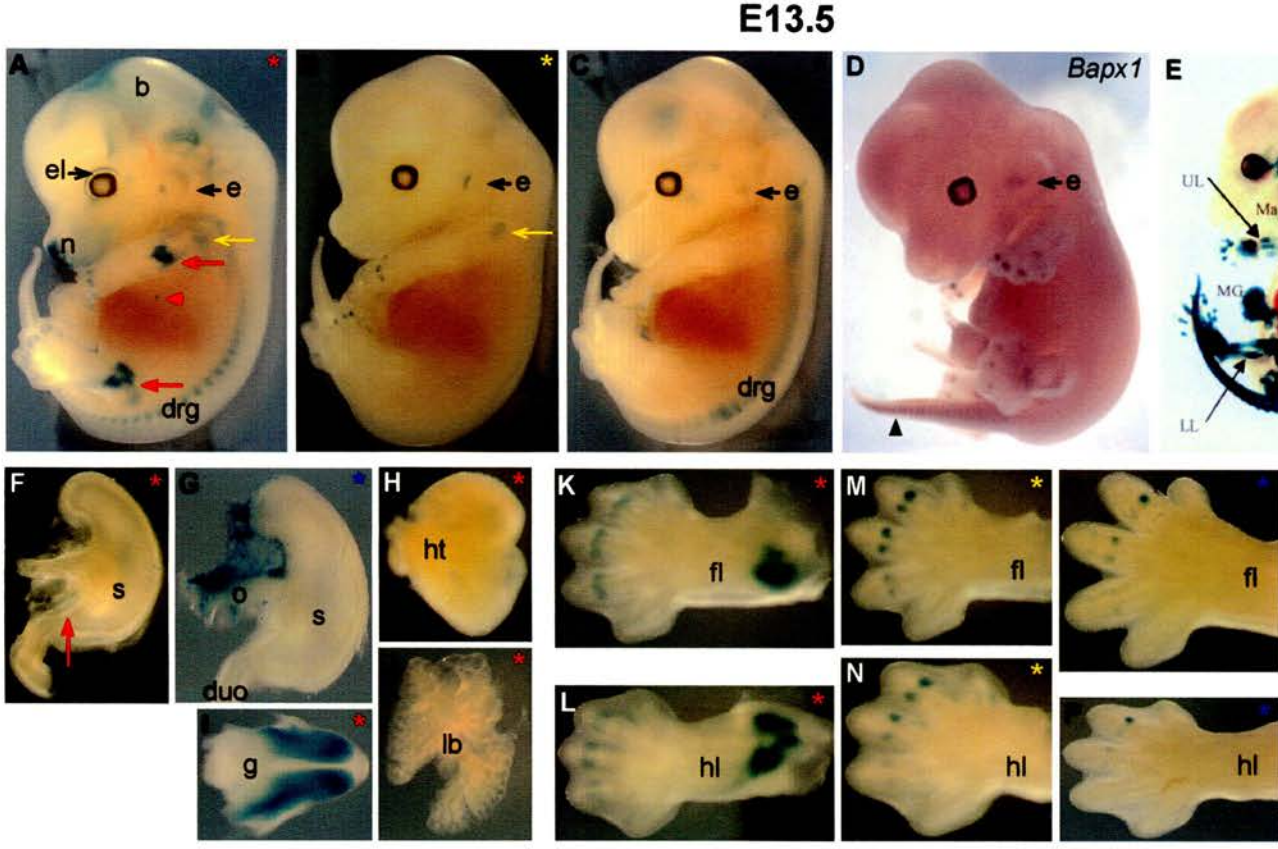


Figure 4.11: Reporter gene expression as directed by DistB1 at E13.5 (**A-C**, **F-P**) in four different transgenic lines, as indicated by different coloured asterisks. Left lateral views of whole embryos (**A-E**) demonstrate some differences between transgenic lines (**A-C**), endogenous *Bapx1* expression (**D**; image provided by Dr. Robert Watson) and the *Bapx1*^{LacZ} mice (**E**; image derives from Akazawa *et al.*, 2000). *LacZ* expression in the digits of lines A266.3 (**A**, **K**, **L**), A266.2 (**B**, **M**, **N**) and A266.4 (**O**, **P**) resembles that of *Bapx1* (**D**). Strong staining in the joints seen in line A266.3 (**A**, **K**, **L**) reflects expression of *LacZ* in the upper and lower limb joints of *Bapx1*^{LacZ} mice (**E**) (red arrows). *LacZ* expression is also detected at a lower level in the humeral joint of the upper limb in lines A266.3 (**A**) and A266.2 (**B**) (yellow arrows). Expression of *LacZ* is detected in the dorsal root ganglia of lines A266.3 (**A**) and A266 (**C**), as opposed to *LacZ* expression in the vertebral column of *Bapx1*^{LacZ} mice (**E**) and *Bapx1* expression in the caudal-most somites (black arrowhead) (**D**). A patch of stain is seen in the ear of two of the transgenic lines (**A**, **B**) which does not, however, correspond to the middle ear domain of *Bapx1* (**D**). In line A266.3, *LacZ* expression is additionally detected in the eyelids and mammary gland primordia (red arrowhead) (**A**). Internally, the reporter gene continues to be expressed in the heart (**H**), genitals (**I**), and lungbuds (**J**). Within the foregut, *LacZ* expression is detected in the same posterior stomach domain of line A266.3 as seen previously at E12.5, albeit at a lower level (vertical red arrow) (**F**). Reporter gene expression along the inner curve of the anterior stomach and proximal oesophagus of line A266.4 (**G**) at E13.5 reflects the pattern seen at E12.5. b, brain; drg, dorsal root ganglia; duo, duodenum; e, ear; el, eyelid; EO, exoccipital bone; fl, forelimb; g, genitals; hl, hindlimb; ht, heart; lb, lungbuds; LL, lower limb; Ma, mandible; Me, mesenteries; MG, midgut; n, nose; o, oesophagus; R, ribs; s, stomach; Sc, scapula; SO, supraoccipital bone; Sp, spleen primordium; Te, temporal bone; UL, upper limb; VC, vertebral column; Zy, zygomatic bone. * Line A266.3; * Line A266.4; * Line A266; * Line A266.2.

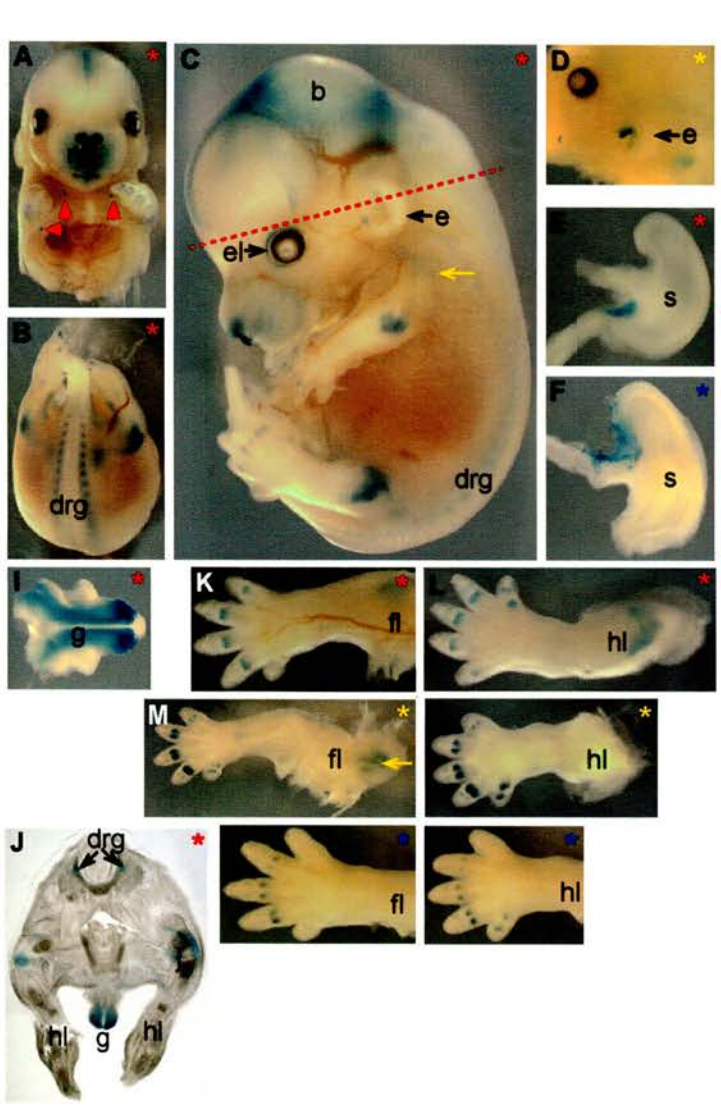


Figure 4.12: Localisation of DistB1-driven reporter gene expression by Xgal staining at E14.5. (**A**) Frontal view, (**B**) posterior view and (**C**) lateral view of an embryo from line A266.3. *LacZ* expression is detected in mammary gland primordia (arrowheads) (**A**) and dorsal root ganglia (**B**, **C**). Transverse sections (**H**, **H'**, **J**) along the plane shown by the dotted red line in **C** confirmed expression in dorsal root ganglia flanking the neural tube. A patch of expression in the ear is visible in lines A266.3 (**C**) and A266.2 (**D**). Transverse sections localise the stain to a domain opposite the pinna of the ear (**G**, **G'**). Two bands of reporter gene expression are observed in the digits of the fore- and hindlimbs of lines A266.3 (**K**, **L**), A266.2 (**M**, **N**) and A266.4 (**O**, **P**). The proximal band is the stronger of the two in all three lines for both fore- and hindlimbs. Prominent *LacZ* expression is observed in the humeral-ulnar joint between the upper and lower forelimb bones of line A266.3 (**C**, **K**) whilst expression in the upper humeral joint of the forelimb is detected at a higher level in line A266.2 (**M**) than in line A266.3 (**C**) (yellow arrows). Prominent *LacZ* expression is also seen in the joint region of the distal part of the femur within the hindlimb of line A266.3 (**B**, **C**, **L**). This joint domain was confirmed by transverse section (**J**). Internally, *LacZ* expression is detected in a portion of the posterior stomach of line A266.3 at E14.5 (**E**), as seen previously at E11.5, E12.5 and E13.5. In line A266.4, foregut expression encompasses the inner curve of the anterior stomach and the proximal oesophagus (**F**), as seen at E12.5 and E13.5. Strong genital staining persists at E14.5 in line A266.3 (**I**, **J**). b, brain; dn, degenerating notochord; drg, dorsal root ganglia; e, ear; el, eyelid; fl, forelimb; g, genitals; hl, hindlimb; j, joint; n, nose; nt, neural tube; p, pinna; s, stomach. * Line A266.3; * Line A266.2; * Line A266.4.

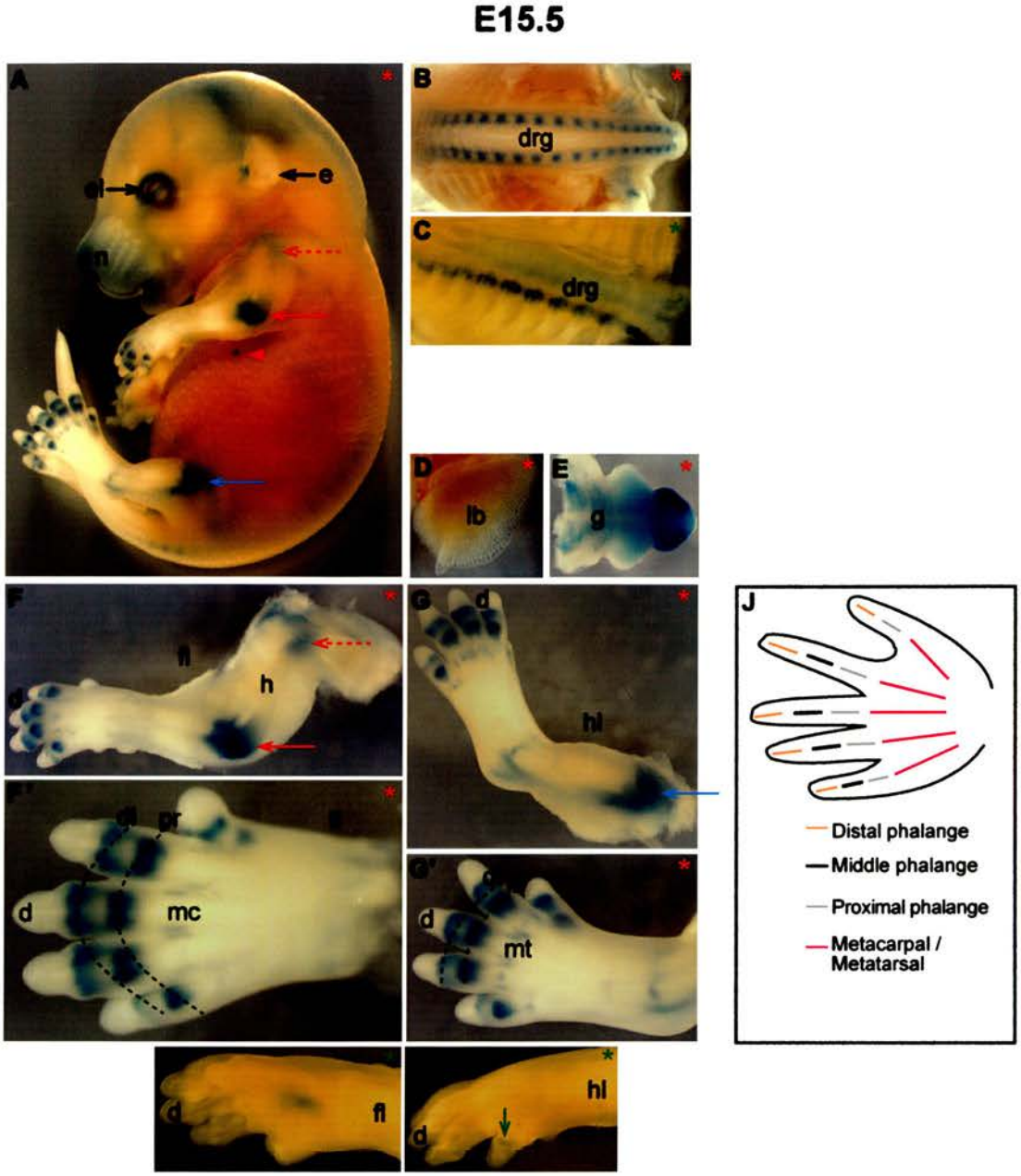


Figure 4.13: Localisation of DistB1-driven *LacZ* expression at later stages (E15.5) of mouse embryogenesis. A lateral view of an embryo from line A266.3 (**A**) demonstrates reporter gene expression in the nose, eyelids, upper and lower limb joints and digits, and mammary gland primordia (red arrowhead). A posterior (**B**) and internal (**C**) view of embryos from lines A266.3 and A266 respectively demonstrate strong *LacZ* expression in the dorsal root ganglia. By E15.5, the only staining seen in internal organs upon dissection is that in the lungbuds (**D**) and genitals (**E**) of line A266.3. Strong expression within the humeral-ulnar joint of the forelimb in line A266.3 is maintained at E15.5 (red arrows) (**A**, **F**), as is high level expression within the distal femur joint of the hindlimb (blue arrows) (**A**, **G**). Weaker expression is observed in the proximal humeral joint of the forelimb (red dotted arrows) (**A**, **F**). The reporter gene is highly expressed within the digits of line A266.3 (**A**, **F**, **F'**, **G**, **G'**), the proximal interphalangeal joint band being stronger than the distal interphalangeal joint band in both fore- (**F'**) and hindlimbs (**G'**). Expression is also visible flanking the distal metacarpal bones of the forelimb (**F'**) and the distal metatarsal bones of the hindlimb (**G'**). A weak domain of expression in the distal forelimb is detected in line A266 at E15.5 (**H**) which differs from the pattern observed in the hindlimb, where a single band of expression is seen in one digit (green arrow) (**I**). (**J**) A cartoon to illustrate the principal bones within the digits and the paw plates of fore- and hindlimbs. d, digits; di, distal; drg, dorsal root ganglia; e, ear; el, eyelid; fl, forelimb; g, genitals; h, humerus; hl, hindlimb; lb, lungbuds; mc, metacarpal; mt, metatarsal; n, nose; pr, proximal. * Line A266.3; * Line A266.

al., 2000). Stain along the entire vertebral column is seen in *Bapx1*^{LacZ} embryos at E13.5 (Figure 4.11E).

The digit expression seen in three transgenic lines at E13.5 is, however, reflective of endogenous *Bapx1* expression (compare Figure 4.11D to 4.11K-P). In both the fore- and hindlimbs of line A266.3, a proximal band of expression surrounding the phalange bones of the digits is visible (Figure 4.11K, L), occupying the same position as staining seen in line A266.2 at E12.5. This digit expression is maintained at E13.5 in line A266.2 (Figure 4.11M, N) and is also apparent in the fore- and hindlimbs of line A266.4 (Figure 4.11O, P). By E14.5, these three lines all exhibit an additional, distal band of expression in the digits (Figure 4.12K-P). The appearance of the proximal expression band prior to the distal band reflects the order of chondrification in the digits: distal phalangeal chondrification lags one day behind that of the proximal phalanges (Rugh, 1990). By E15.5, the digit pattern is only maintained at a high level in line A266.3. In addition to strong *LacZ* expression corresponding to cartilaginous condensations in the digits, both fore- and hindlimbs reveal additional domains of expression flanking the metacarpal (forelimb) and metatarsal (hindlimb) bones (Figure 4.13F', G'). Sagittal sections of A266.3 limbs at E15.5 permitted closer examination of the distribution and intensity of reporter gene expression within the digits. The three expression domains appear to correspond to the positions of the proximal and distal interphalangeal joints and the metacarpo-/metatarsophalangeal joints (Figure 4.14). The two principal (proximal and distal) bands of expression in the digits are connected by a narrow band of expression along the side of each digit (Figure 4.14B, G, H). The staining pattern is shown to be a continuous ring of expression around the bone by transverse sections (Figure 4.14C).

Furthermore, although *Bapx1* expression is not detected in the limb joints of the long bones by wholemount *in situ* hybridisation, the upper and lower limb joints exhibit strong *LacZ* expression in E13.5 *Bapx1*^{LacZ} mice (Figure 4.11E) which is comparable to that seen in line A266.3 at E13.5 (Figure 4.11A). Expression in the upper forelimb joint is also present at this stage, albeit weaker, in line A266.2 (Figure 4.11B). Transverse sections of line A266.3 at E14.5 confirm that the region of strong *LacZ* expression in the hindlimb corresponds to the joint at the distal part of the femur which forms a connection to the tibia/fibula (Figure 4.12J). This is further corroborated by sagittal sections taken through E15.5 hindlimbs (Figure 4.14D-F),

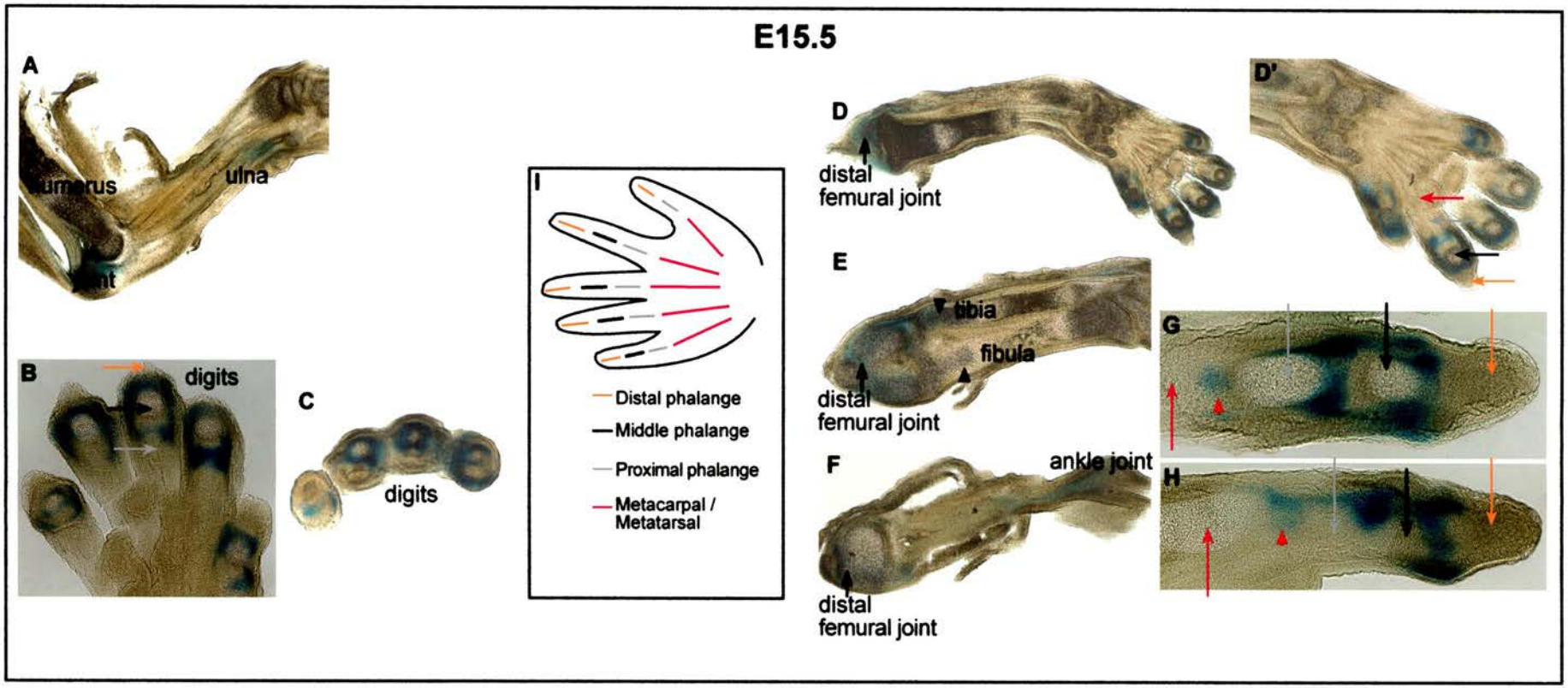


Figure 4.14: DistB1-directed *LacZ* expression within the forelimbs (**A-C**) and hindlimbs (**D-H**) of an E15.5 embryo from line A266.3, as seen in sagittal (**A**, **B**, **D-H**) and transverse (**C**) sections. (**I**) A cartoon to illustrate the principal bones within the digits and the paw plates of fore- and hindlimbs. *LacZ* expression is detected in the ulna of the distal forelimb and the humeral-ulnar joint by sagittal section (**A**). Strong expression is seen corresponding to the regions of the interphalangeal joints of the forelimb digits (**B**). Transverse sections demonstrate that the staining represents a continuous ring of expression around the digit bones (**C**). Expression of the reporter gene is detected in the distal femural joint of the hindlimb (**D**, **E**, **F**) and also within the proximal portion of the tibia and fibula (black arrowheads) (**E**). Staining in the vicinity of the ankle joint is also visible (**F**). Strong *LacZ* expression is detected in regions corresponding to the interphalangeal joints of the hindlimb digits (**D** and close up in **D'**, **G**, **H**). Weaker *LacZ* expression is also visible in the metatarsophalangeal joints of the hindlimb (red arrowheads) (**G**, **H**). Coloured arrows denote the digit bones as outlined in **I**.

which additionally demonstrate reporter gene expression in the ankle joint (Figure 4.14F) that is also visible in whole E15.5 embryos (Figure 4.13G, G'). Forelimb sagittal sections at E15.5 illustrate *LacZ* expression in the joint connecting the humerus and the ulna (Figure 4.14A). In addition to expression in the joints, sagittal sections of E15.5 limbs reveal some expression in the bones of the ulna (forelimb) and tibia and fibula (hindlimb) similar to that detected in *Bapx1*^{LacZ} mice at E13.5 (Figure 4.14A, E).

Dissection of internal organs at E13.5 reveals that *LacZ* expression in the heart, lungs, posterior stomach and genitals is maintained in line A266.3 (Figure 4.11F, H-J). Line A266.4 again reveals *LacZ* expression along the inner curve of the anterior stomach and proximal oesophagus, as at E12.5 (Figure 4.11G). These internal expression patterns are maintained at E14.5, excepting heart and lung expression in line A266.3 which is not detected (Figure 4.12E, F, I). Lung expression is, however, detectable at E15.5, along with extremely strong genital staining, although staining is no longer visible in the posterior stomach of either line A266.3 or A266.4 (Figure 4.13D, E). Table 4.2 summarises the tissues in which DistB1-directed *LacZ* expression is observed in the four expressing DistB1 transgenic lines.

Table 4.3 details, for each ProxB and DistB1 transgenic line, the developmental stages at which reporter gene expression is detected. Precise litter sizes and numbers of embryos expressing the transgene are tabulated in Appendix 3.

4.2.6 Investigation of loci in the *Bapx1* region other than *Bapx1* which may be influenced by ProxB and DistB1

The data presented above demonstrates enhancer activity of the ProxB and DistB1 CNSs which shows similarity, but not perfect identity, to the reported *Bapx1* expression pattern. Consequently, investigation of other genes which may be regulated by the ProxB or DistB1 elements was undertaken. Given the degree of conservation of these elements between human and mouse, in addition to the conservation of ProxB to zebrafish, it was considered extremely likely that the elements influenced a gene common to the species in which they are preserved. The chromosomal region encompassing in excess of 1Mb DNA either side of *Bapx1* is

External	Tissue and corresponding **Figure** reference														
Line designation	Midbrain	Hindbrain	Dorsal root ganglia	Forelimb joints	Hindlimb joints	Forelimb bones	Hindlimb bones	Digits	Ear		Nose		Eyelid	Mammary gland primordia
									Early	Late	Early	Late		
A266			**4.11**C; **4.13**C					**4.13**I						
A266.2				**4.11**B; **4.12**M				**4.10**U; **4.11**B, M, N; **4.12**M, N		**4.11**B; **4.12**D				
A266.3	**4.9**B, C, F, G, H; **4.10**A, H; **4.11**A	**4.9**B, C, F, H; **4.10**A, H; **4.11**A	**4.10**H, Q, R; **4.11**A; **4.12**B, H, J; **4.13**B	**4.9**D*, I*, K*; **4.10**A, H, P; **4.11**A, K; **4.12**C, K; **4.13**A, F; **4.14**A	**4.10**H; **4.11**A; **4.12**C, J, L; **4.13**A, G; **4.14**D, E, F	**4.10**H, S; **4.11**A; **4.14**A	**4.10**H, T; **4.11**A, L; **4.13**G; **4.14**E	**4.11**A, K, L; **4.12**C, K, L; **4.13**A, F, G; **4.14**A, B, C, D, G, H	**4.9**B, F; **4.10**A, H	**4.11**A; **4.12**C, G; **4.13**A	**4.9**F; **4.10**A, H, P	**4.11**A; **4.12**A; **4.13**A	**4.11**A; **4.12**C; **4.13**A	**4.11**A; **4.12**A
A266.4								**4.11**O, P; **4.12**O, P						

Internal organs	Tissue and corresponding **Figure** reference				
Line designation	Stomach		Heart	Lungbuds	Genitals
	Anterior	Posterior			
A266		**4.10**F, G			
A266.2					
A266.3		**4.10**D, M; **4.11**F; **4.12**E	**4.9**A, B, D, F, I; **4.10**A, J; **4.11**H	**4.9**E, J, K; **4.10**D, K, Q; **4.11**J; **4.13**D	**4.10**L; **4.11**I; **4.12**I, J; **4.13**E
A266.4	**4.10**N; **4.11**G	**4.10**E; **4.12**F			

Table 4.2: A summary of tissues in which DistB1-directed reporter gene expression is observed in the four expressing DistB1 transgenic lines. * Presumed region of expression as deduced from pattern at later developmental stages.

Line designation	E8.5	E9.5	E10.5	E11.5	E12.5	E13.5	E14.5	E15.5	E16.5
A266	–	–	X	✓	✓	✓	✓	✓	–
A266.1	–	–	X	X	X	X	X	–	–
A266.2	X	–	X	X	✓	✓	✓	X	–
A266.3	✓	✓	✓	✓	✓	✓	✓	✓	–
A266.4	–	–	X	✓	✓	✓	✓	X	–
A266.5	–	–	X	–	X	X	X	–	–
A267	–	✓	✓	✓	✓	✓	✓	–	–
~~A269~~									
A269.1	–	X	X	X	X	X	X	X	–
A269.2	–	✓	✓	✓	✓	✓	✓	✓	✓
A269.3	–	–	X	X	X	X	X	–	–
~~A269.4~~									
A269.5	–	✓	✓	✓	✓	✓	✓	✓	–
~~A269.6~~									

Table 4.3: A table to show the detection (✓) or absence (X) of reporter gene expression for the ProxB and DistB1 transgenic lines during embryogenesis. DistB1 lines appear above the horizontal red division whilst ProxB lines are below it. Developmental stages not tested are indicated by –. Line A269 did not transmit the transgene. Lines A269.4 and A269.6 could not be assayed due to failure to breed and infertility respectively.

sparsely populated by genes in mouse and human (see Chapter 3). Within 3Mb of genomic sequence centred on *Bapx1*, Ensembl reveals only the *Bapx1* cluster to be shared between human chromosome 4p and mouse chromosome 5. The cluster comprises *NM_148894*, *Q9P2L9*, *BAPX1* and *RAB28*. *NM_148894* and *BAPX1* are also found in a syntenic region of zebrafish, with the possibility that additional genomic sequence and an updated assembly in this species will reveal the presence of *RAB28*. It was thus reasonable to pursue investigation of ProxB and DistB1 as control elements for a gene in this locality.

4.2.6.1 The existence of a noncoding RNA?

Subsequent to the initiation of the transgenic bioassay, but prior to amassing the expression data, a former version of Human Ensembl (March 2003) predicted the existence of a "novel" gene lying between *BAPX1* and *RAB28* in human genomic DNA only (Figure 4.15). It was predicted to comprise four exons and the direction of transcription was the same as for immediate flanking genes. Mapping relative to ProxB and DistB demonstrated that the "novel" gene lay between *BAPX1* and DistB, with ProxB positioned in the first intron (Figure 4.16). Ensembl BLAST analysis of the predicted transcript returned a single match to a human cDNA (Accession number: AK096368) derived from NT2 neuronal precursor cells after 5 weeks' retinoic acid induction. The predicted exons of the "novel" gene demonstrated perfect identity to the human cDNA. Interestingly, the ORF predicted by Ensembl is contained entirely within exon4 however closer analysis reveals the presence of one too many nucleotides (a 442bp coding region) between the predicted ATG start and TGA termination codons. This may partly explain the subsequent removal of this "novel" gene from Human Ensembl.

Neither Ensembl nor NCBI BLAST analyses reveal any homologous genes corresponding to the human cDNA which, together with the results mentioned above concerning the apparent absence of a genuine ORF, raises the possibility that the "novel" gene may represent a noncoding RNA (ncRNA). The latter term is used to describe transcripts that lack an apparent ORF and do not encode a protein product. If the latter were true, it would relieve selectional pressure at the "novel" gene locus thus accounting for its high sequence divergence in the equivalent mouse region in

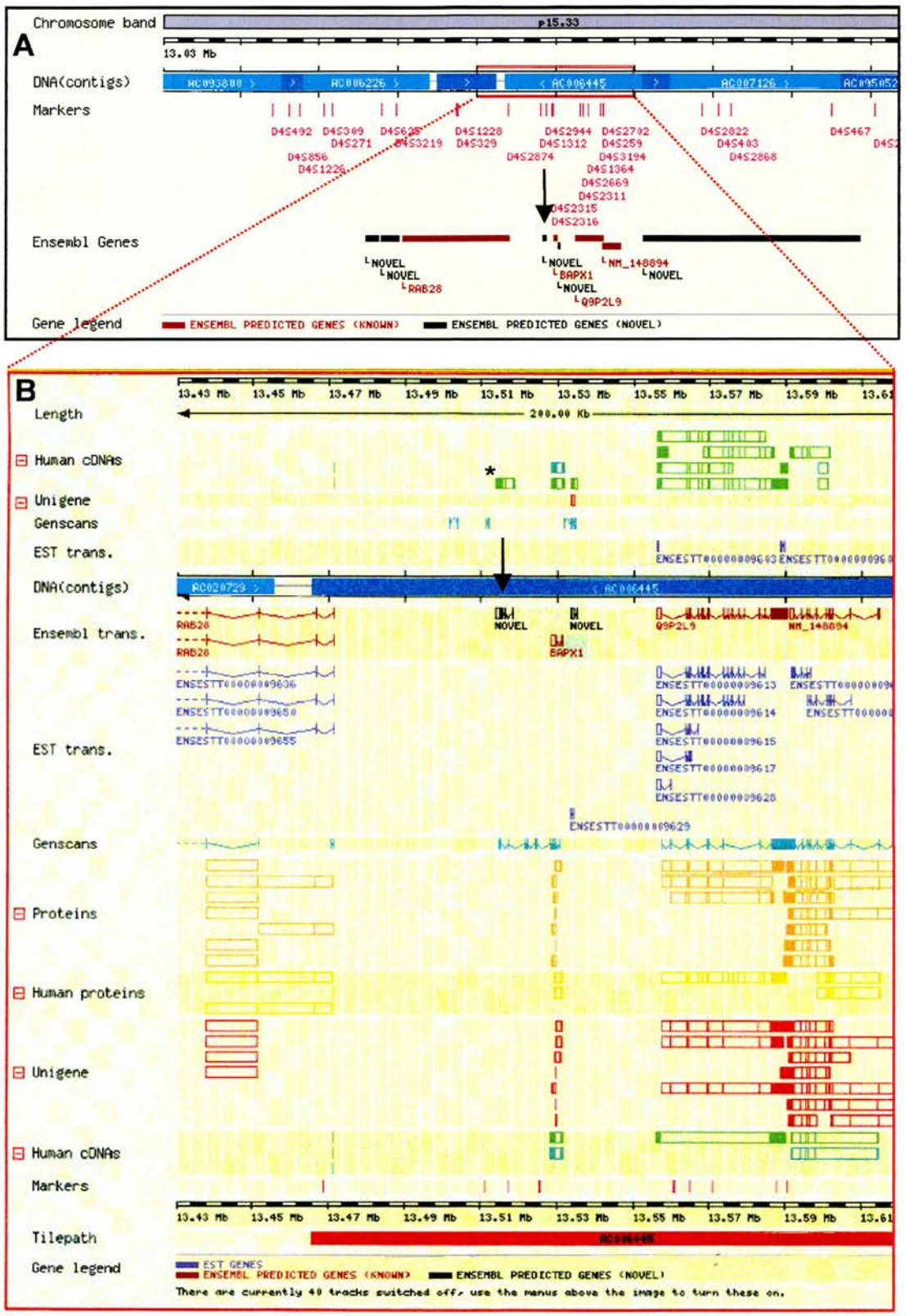


Figure 4.15: **(A, B)** Identification of a novel gene lying between *BAPX1* and *RAB28* in Human Ensembl. *BAPX1* is highlighted in green and the novel gene is indicated by a black arrow. The human cDNA sharing sequence identity with the novel gene is indicated by a black star in **B**.

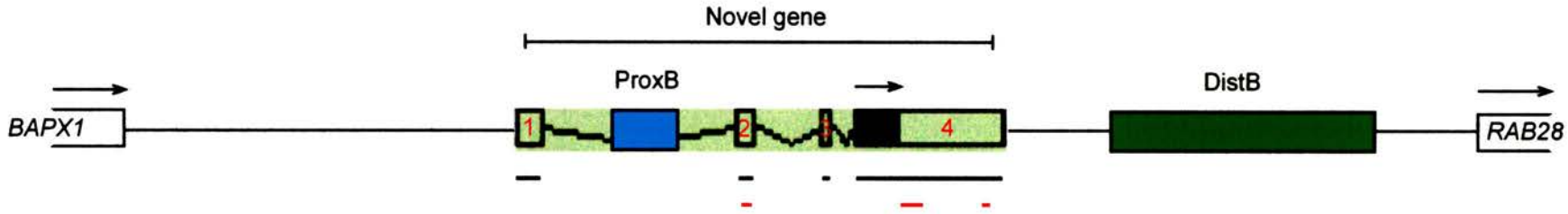


Figure 4.16: A schematic to illustrate the relative positions of ProxB and DistB with respect to the novel gene prediction in human genomic DNA sequence. Exons of the novel gene are boxed and numbered with the predicted open reading frame shaded black. The direction of transcription is indicated by black arrows. The horizontal black lines beneath the predicted exons identify sequence showing nucleotide identity to a human cDNA. The horizontal red lines indicate the regions of human genomic DNA sequence that are conserved in the equivalent genomic region of mouse.

which only short stretches of sequence corresponding to exons 2 and 4 are aligned by BLAST (see Figure 4.16). Interestingly, the Ensembl Genscan programme predicts a two-exon transcript flanking ProxB in zebrafish (see Figure 3.7). BLAST analyses did not reveal any significant sequence identity to either the "novel" gene within the human genome or the equivalent region of mouse. Notably, ncRNAs are expressed in a tissue-specific manner (Erdmann *et al.*, 2001a), thus the possibility arises that ProxB and/or DistB1 are control elements which drive tissue-specific expression of a putative ncRNA positioned between *Bapx1* and *RAB28*.

4.2.6.2 Examination of *RAB28* and the putative ncRNA

Due to the identification of the "novel" gene in human but not mouse, it was considered advantageous to determine whether the "novel" gene was also transcribed in mouse and, if so, where it was expressed. These questions were addressed by RT-PCR and *in situ* hybridisation. *RAB28* was investigated in parallel since, although the expression data available for *RAB28*, *NM_148894* and *Q9P2L9* was suggestive of ubiquitously expressed genes (see Chapter 3), *RAB28* lay immediately downstream of DistB and ProxB which may therefore be upstream *cis*-acting regulators. The expression data reports the existence of two *RAB28* isoforms, one of which is predominantly found in the testis whilst the other is ubiquitous (Brauers *et al.*, 1996). Yet extensive literature searches failed to unearth any *RAB28 in situ* data thus information relating to spatial and temporal expression patterns was limited to matches found in cDNA libraries and corresponding to random ESTs.

Following discovery of both ProxB and DistB1 enhancer activity at early developmental stages, it was decided to investigate expression of *RAB28* and the "novel" gene at E10.5. Accordingly, total RNA was isolated from E10.5 mouse embryo heads and bodies as described in Materials and Methods. The RNA was used to generate two sources (head and body) of cDNA by reverse transcription (RT). RT-PCR primers corresponding to mouse sequence were designed to span intron-exon boundaries in order to distinguish between spliced and unspliced cDNA by the size of the PCR product. The primer locations and expected PCR products, along with the RT-PCR results, are shown in Figure 4.17.

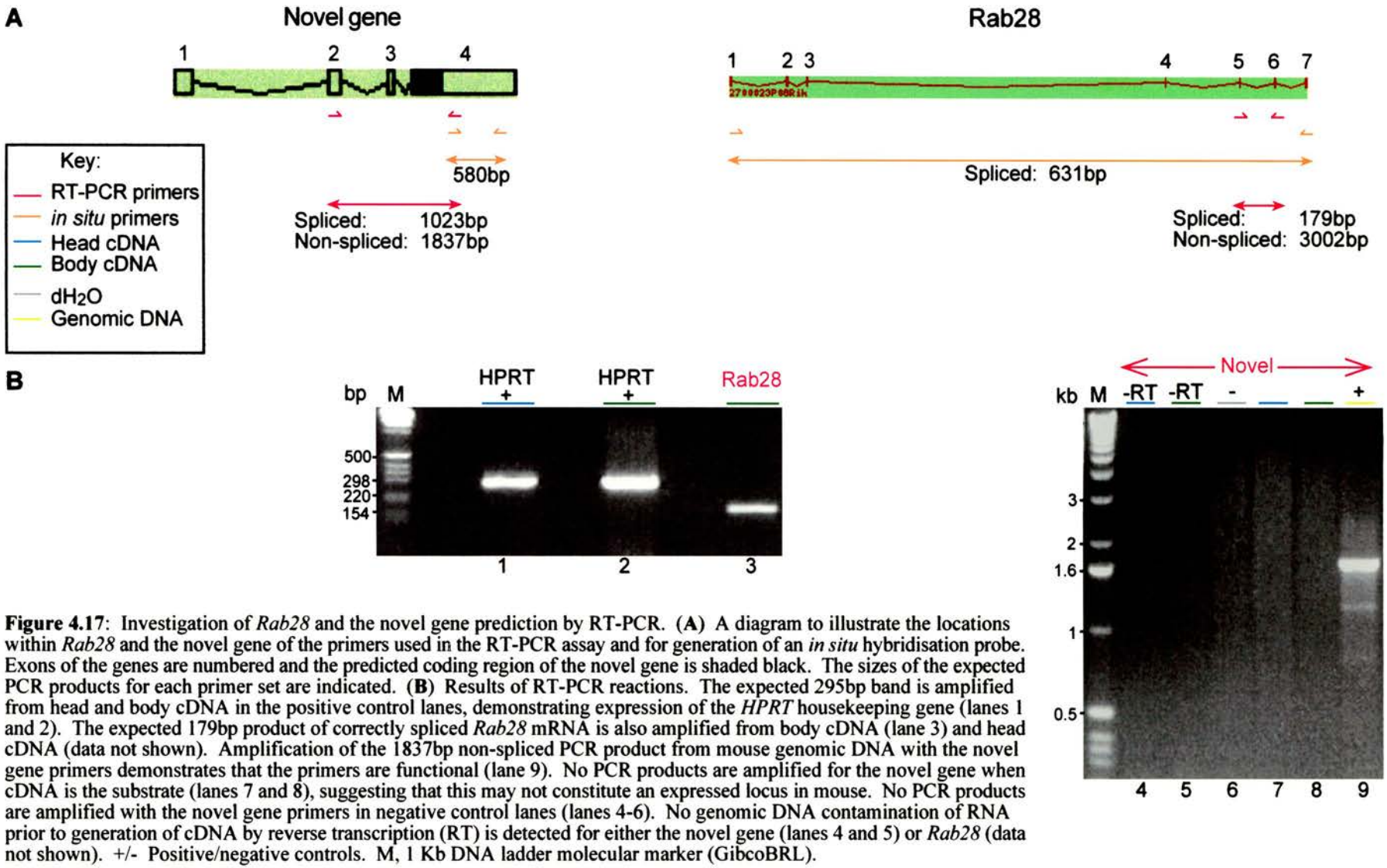


Figure 4.17: Investigation of *Rab28* and the novel gene prediction by RT-PCR. (**A**) A diagram to illustrate the locations within *Rab28* and the novel gene of the primers used in the RT-PCR assay and for generation of an *in situ* hybridisation probe. Exons of the genes are numbered and the predicted coding region of the novel gene is shaded black. The sizes of the expected PCR products for each primer set are indicated. (**B**) Results of RT-PCR reactions. The expected 295bp band is amplified from head and body cDNA in the positive control lanes, demonstrating expression of the *HPRT* housekeeping gene (lanes 1 and 2). The expected 179bp product of correctly spliced *Rab28* mRNA is also amplified from body cDNA (lane 3) and head cDNA (data not shown). Amplification of the 1837bp non-spliced PCR product from mouse genomic DNA with the novel gene primers demonstrates that the primers are functional (lane 9). No PCR products are amplified for the novel gene when cDNA is the substrate (lanes 7 and 8), suggesting that this may not constitute an expressed locus in mouse. No PCR products are amplified with the novel gene primers in negative control lanes (lanes 4-6). No genomic DNA contamination of RNA prior to generation of cDNA by reverse transcription (RT) is detected for either the novel gene (lanes 4 and 5) or *Rab28* (data not shown). +/- Positive/negative controls. M, 1 Kb DNA ladder molecular marker (GibcoBRL).

The *HPRT* (*Hypoxanthine Guanine Phosphoribosyl Transferase*) gene is a ubiquitously expressed 'housekeeping gene' that served as a positive control.

RAB28 primers amplified the expected 179bp band from both body and head cDNA confirming that, as expected, the gene is both transcribed and correctly spliced in the head and body. The failure to amplify a PCR product from genomic DNA is likely to result from experimental difficulties in amplifying the expected 3kb band. The amplification of a 1.8kb non-spliced "novel" gene band from genomic DNA demonstrates that the "novel" gene primers are functional (Figure 4.17B). However, the absence of a 1kb correctly spliced band amplified from cDNA suggests that the "novel" gene, at E10.5 at least, is not transcribed or spliced in mouse.

PCR primers (including one containing a T7 promoter tag for subsequent DIG-labelling) were designed to the mouse homologues of both *RAB28* and the "novel" gene for *in situ* hybridisation purposes. The *RAB28* primers spanned six introns therefore would only amplify correctly spliced mRNA present in the total RNA mix from E10.5 head/body. "Novel" gene PCR primers were designed to the mouse equivalent of the predicted exon4 due to the location of two conveniently spaced, short conserved regions in this exon that constituted the principal sequence conserved to human at the mouse genomic "novel" gene locus (see Figures 4.16 and 4.17). The expected PCR products were successfully amplified and wholemount *in situ* hybridisation of wildtype CD1 E9.5 and E10.5 mouse embryos was performed. No differences were observed, however, between embryos hybridised with either *RAB28* or the "novel" gene probes; specific staining patterns were not seen and after 5 hours in staining solution, the reactions were terminated owing to the appearance of the embryos being indicative of background stain with trapping of the probe in each case in the head. The inability to detect *RAB28* transcripts by *in situ* hybridisation is in contrast to the demonstration of *RAB28* expression by RT-PCR, together with the existence of multiple *RAB28* ESTs. These results therefore suggest that *RAB28* is either expressed at a sufficiently low level as to escape detection by wholemount *in situ* hybridisation, instead requiring the higher sensitivity of PCR to detect its presence, or the RNA *in situ* probe is not functioning correctly. In the absence of a positive control, the "novel" gene *in situ* experiments are inconclusive however RT-PCR data suggests that this "novel" gene locus is not active at E10.5 in

mouse. Considering the results of the transgenic bioassay at early developmental stages, it would thus appear that the "novel" gene locus has no connection to the tissue-specific expression directed by ProxB or DistB1.

4.3 Conclusions and Discussion

4.3.1 ProxB and DistB1 function as tissue-specific enhancer elements

My data provides evidence that both the ProxB and DistB1 CNSs are regulatory elements which exert an influence on gene expression from the early stages of embryogenesis. The fact that two of the five ProxB and two of the six DistB1 lines tested did not demonstrate reporter gene expression is likely to be due to their genomic integration sites. Indeed the transgenic construct into which the elements were cloned is reported to have a relatively tight promoter such that a somewhat lower percentage of integration sites will express the transgene (Yee and Rigby, 1993). This in turn should enhance detection of *bona fide* regulatory activity whilst limiting ectopic reporter gene expression.

Enhancers have long been known to function in an orientation- and distance-independent manner with regard to their cognate genes (Serfling *et al.*, 1985; Lewin, 2000). The coding sequences in the vicinity of the ProxB and DistB1 elements are all transcribed in the same direction and with this in mind, ProxB was cloned into the reporter construct in the same orientation as *LacZ*, whilst DistB1 was cloned in the opposite orientation. Both direct *LacZ* expression in a temporally- and spatially-regulated manner, alluding to their being genuine enhancers. Furthermore, ProxB and DistB1 independently drive reporter gene expression in specific, non-overlapping domains thus each may target a separate gene, or may constitute two tissue-specific enhancers of a single gene. Amalgamating the data presented in both this, and the previous chapter, I favour the latter proposal and moreover, believe that *Bapx1* may constitute their target gene. The following discussion explores various lines of evidence that support this theory.

4.3.2 Sequences other than *Bapx1* potentially influenced by ProxB and/or DistB1

The operation of regulatory elements over enormous genomic distances has recently been reported, for example in the identification of a long range *Shh* enhancer acting over 1Mb (Lettice *et al.*, 2003). It is equally accepted that the ProxB and DistB1 elements, having demonstrated their enhancer activity when cloned directly upstream of the *LacZ* reporter gene, may function *in vivo* over a significant distance. However, as stated earlier, there are very few genes either known or predicted to lie within the 3Mb genomic region centred around *Bapx1* (and hence the ProxB and DistB1 elements) for both mice and humans. In humans, for which the finished genomic sequence is available, only the four genes clustered around and including *BAPX1* (*RAB28*, *BAPX1*, *Q9P2L9* and *NM_148894*) reside within this 3Mb expanse. These all appear to have homologues in mouse and synteny is conserved to fish (see Chapter 3). It is therefore reasonable to assume that ProxB and DistB1 influence the expression of one of these genes.

Protein products of the *RAB28*, *Q9P2L9* and *NM_148894* genes have been predicted but the function of these genes is currently undetermined. The expression data available in each case however is suggestive of ubiquitously expressed genes; transcripts from each derive from a wide variety of cDNA tissue library sources over equally widespread developmental stages. The tissue-specific expression driven by ProxB and DistB1 is not consistent with the staining pattern expected of these seemingly ubiquitous genes. Nevertheless, the transgenic DistB1 construct comprises only part of an 8kb conserved region between mouse and human thus it is possible that the entire region can account for a broad expression pattern. If this was true, perhaps it would be more likely to affect local expression of the gene immediately downstream (*RAB28*) since ubiquitously expressed genes tend to have general "housekeeping" functions within cells thus are unlikely to require distantly located enhancers to direct expression in a specific spatial or temporal manner. However, there exists a *Fugu* homologue of *RAB28*, yet there is no sign of an equivalent DistB1 regulatory element in fish.

4.3.2.1 What of the "novel" gene?

The "novel" gene formerly identified by human Ensembl as residing between *BAPX1* and *RAB28* does not appear to constitute the target of the ProxB or DistB1 regulatory elements. RT-PCR experiments based on the equivalent region in mouse do not detect the existence of a spliced mRNA product at a developmental stage at which ProxB and DistB1 are active. Nor are transcripts detected by wholemount *in situ* hybridisation. The "novel" gene locus does, however, bear the hallmarks of a noncodingRNA (ncRNA). A human EST demonstrates complete identity to the predicted exons of the "novel" gene, but no protein products corresponding to the EST have been identified. These observations indicate that this locus is transcribed and correctly spliced in humans, but that the mRNA is not translated.

Numerous short ncRNAs such as ribosomal RNA, transfer RNA and small interfering RNAs (siRNAs) have been identified and characterised and they form a distinct class from the longer ncRNAs which are processed in the same way as mRNA, hence occasionally referred to as mRNA-like ncRNAs (Erdmann *et al.*, 2000). Examples of longer ncRNAs involved in gene regulation include the X-chromosome inactivator *Xist*, which functions early during mammalian female development to effect dosage compensation (Brown *et al.*, 1991; Borsani *et al.*, 1991; Brockdorff *et al.*, 1991; Brown *et al.*, 1992; reviewed by Avner and Heard, 2001), and *H19*, which is transcribed from an imprinted locus and for which no precise role has yet been assigned although the data is consistent with the ncRNA being functional (Hurst and Smith, 1999; reviewed by Thorvaldsen and Bartolomei, 2000; Erdmann *et al.*, 2001b; Szymanski *et al.*, 2003). ncRNAs have been identified in plants and animals however they are not generally well conserved between mammalian species which, together with the difficulties associated with defining exon boundaries, renders their identification difficult. The latter could thus help to account for the removal of the "novel" gene prediction from human Ensembl, a browser that, at present, is principally most effective in the identification of putative *coding* sequences. Furthermore, the limited human/mouse conservation at the "novel" gene locus is consistent with the presumed lack of extensive sequence constraint at a *noncoding* RNA locus. It is noteworthy that the identification of 4280

candidate ncRNA transcripts amongst the extensive RIKEN mouse full-length cDNA collection has recently been reported (Numata *et al.*, 2003). This serves to highlight their emerging importance as a major functional subclass of processed transcripts in mammals. Whether the predicted transcript in the vicinity of zebrafish ProxB has any connection to the human EST spliced from human ProxB-flanking sequence is an intriguing prospect. The function of these putative ncRNAs is unknown and, considering the conservation of ProxB to zebrafish, it is tempting to speculate that zebrafish ProxB is also an active element but that, extrapolating from the data obtained in mouse, it does not regulate the expression of the putative zebrafish transcript.

4.3.3 Could ProxB and DistB1 constitute tissue-specific enhancers of *Bapx1*?

4.3.3.1 Postulation of ProxB as an evolutionarily conserved enhancer of *Bapx1* craniofacial expression

ProxB drives *LacZ* expression in a highly tissue-specific manner, being detected initially in the inferior portion of the first branchial arch during early mouse development. *Bapx1* is similarly detected in this region by E10.5 with the more discrete expression domain corresponding to the highest intensity of *LacZ* expression in ProxB transgenic mice. The slightly broader ProxB domain could be explained by the appreciably higher sensitivity of *LacZ* detection as compared to *in situ* hybridisation. Expanding the significance of this ProxB domain, *Bapx1* is not only expressed in the first branchial arch of mammalian vertebrates, but has been detected in the homologous regions of *Danio rerio* (zebrafish) (Miller *et al.*, 2003) and *Xenopus laevis* (frog) (Newman *et al.*, 1997), alluding to a continual role for *Bapx1* in craniofacial evolution. Notably, the ProxB element is conserved in zebrafish which, together with the comparable region of expression, provides a preliminary indication that it may constitute an evolutionarily conserved regulator of *Bapx1*.

The endogenous *Bapx1* expression domain during early vertebrate development corresponds to the precursor of Meckel's cartilage and this association

between *Bapx1*-expressing cells and the Meckel's cartilage region persists. As branchial arch derivatives develop in mice, *Bapx1* expression is quantitatively reduced in this region but qualitatively still detected at E12.5 in a domain associated with Meckel's cartilage and in mesenchymal condensations that ultimately give rise to structures of the middle ear (Tucker *et al.*, 2003, submitted). Expression directed by ProxB at E12.5 similarly appears to encompass the region of the future middle ear yet strong *LacZ* staining is also seen along the precartilage primordium of Meckel's cartilage. By later developmental stages, *Bapx1* is restricted to the middle ear where it is essential for normal patterning of the middle ear-associated bones (Tucker *et al.*, 2003, submitted). *LacZ* staining in ProxB mice is most pronounced in the ear, but reporter gene expression is also strongly maintained extending along Meckel's cartilage and by E13.5 is also detected in the submandibular (salivary) glands. Interestingly, $Bapx1^{LacZ}$ mice also exhibit expression in the mandible (lower jaw) similar to that seen for ProxB (Akazawa *et al.*, 2000). Thus, aside from the salivary gland expression, reporter gene expression driven by ProxB closely mimics documented sites of *Bapx1* expression, in particular the association with cartilaginous condensations.

Interestingly, expression of *Nkx3.1*, another member of the NK-2 family of homeodomain-containing transcription factors, is also reported in mouse salivary glands (Schneider *et al.*, 2000). It is therefore possible that *Bapx1* (*Nkx3.2*) is also expressed in salivary glands, but at such a low level as to escape detection by conventional means. *Bapx1* need not necessarily play a vital role in this tissue if that function was accomplished by its paralogue *Nkx3.1*. Overlapping expression of *Nkx3.1* and *Bapx1* (*Nkx3.2*) has previously been reported in the sclerotome of the axial skeleton and indeed a role for *Nkx3.1* in skeletal patterning was only revealed by the enhanced skeletal defects of *Nkx3.1*/*Nkx3.2* double knockout mice as compared to *Nkx3.2* homozygous mutants alone (Herbrand *et al.*, 2002). Whether *Bapx1* is expressed endogenously in the salivary glands as opposed to this being a region of ectopic ProxB expression thus remains to be determined.

Comparison of *Bapx1* expression in evolutionarily diverged species provides further support for ProxB being a conserved regulator of *Bapx1*. Importantly, there is an evolutionary relationship between the middle ear bones of mammals and the

jaw bones of non-mammalian vertebrates such as fish. In non-mammalian jawed vertebrates the bones homologous to the mammalian middle ear ossicles comprise the proximal jaw bones that form the jaw articulation (primary jaw joint). In zebrafish, *Bapx1*-expressing cells are detected surrounding the posterior end of the lower jaw and are also reported in the posterior end of Meckel's cartilage (Miller *et al.*, 2003). In this species, *Bapx1* was shown to be required for formation of the jaw joint, the joint-associated retroarticular process of Meckel's cartilage and the retroarticular bone (Miller *et al.*, 2003). Expression of the *Xenopus* homologue of *Bapx1*, *Xbap*, is also detected in a region encompassing the jaw joint (Newman *et al.*, 1997) and in chick (*Gallus gallus*), *Bapx1* is detected within Meckel's cartilage and the jaw joint region (Tucker *et al.*, 2003, submitted). The principal difference between *Bapx1* and ProxB expression patterns thus appears to be that *Bapx1* is downregulated in cartilage prior to the ossification of bones but maintained in the jaw joint/middle ear where it has critical patterning roles. By contrast, *LacZ* staining in ProxB mice is not downregulated, but instead persists in cartilage throughout the later stages of development at which point it ultimately marks the fusion of the left and right Meckel's cartilage at the distal end of the lower jaw.

The differences between endogenous *Bapx1* expression and the *LacZ* staining pattern may be explained in part by the increased sensitivity of *LacZ* detection as opposed to mRNA detection by *in situ* hybridisation. In this respect, it is notable that *LacZ* staining of ProxB mice overtly appears to reflect *LacZ* staining in the comparable region of $Bapx1^{LacZ}$ mice. Considering that *Bapx1* has been demonstrated to be essential for formation of the jaw joint in fish and for normal middle ear development in the mouse, its expression in Meckel's cartilage, once these structures have been specified, may be superfluous to developmental requirements. Perhaps, therefore, there exist trans-acting regulatory proteins that associate either with Bapx1 or with other *in cis Bapx1* regulatory elements to limit its expression to those domains in which it is essential to normal development. In either case, it would escape detection in both the ProxB and the $Bapx1^{LacZ}$ transgenic mice. Notably, the correct localisation of *Bapx1* expression in the jaw joint of zebrafish is reportedly dependent in part on repression by *hand2* (Miller *et al.*, 2003). It is also possible that perdurance of *LacZ*, at the mRNA or protein level, accounts for the

ability to detect *LacZ* in Meckel's cartilage at later stages of embryogenesis; *Bapx1* transcripts and/or the Bapx1 protein may be less stable.

From an evolutionary perspective, *Bapx1* has acquired multiple developmental roles in vertebrates supplementary to its ancient function in the gut discovered in the invertebrate *Drosophila* (Azpiazu and Frasch, 1993). Of its importance with regard to the specification of skeletal elements, the conserved role of *Bapx1* in the jaw joint/middle ear region is particularly fascinating given the morphological transformation this region has undergone during vertebrate evolution. This observation points to the existence of a conserved genetic network in vertebrates directing *Bapx1* expression to the craniofacial skeleton. The parallels between the *Bapx1* expression pattern and reporter gene expression driven by the ProxB enhancer, together with conservation of the latter to zebrafish, provide strong evidence that ProxB may constitute a *Bapx1* regulatory element and thus be an integral part of this craniofacial genetic network. This proposal would be greatly strengthened by the discovery of the ProxB element in *Xenopus* and chick, for which sufficient sequence is not currently available. Furthermore, it would be interesting to investigate the role of *Bapx1* in the lamprey (non-jawed vertebrate) and to learn whether these ancient ancestors also possess a ProxB element. The results of such a study would likely shed a great deal more light on various aspects of craniofacial evolution.

An additional experiment initiated to explore the putative regulatory relationship between ProxB and *Bapx1* was the breeding of line A267 (ProxB) mice into the *Bapx1*-/- genetic background. This study was in progress at the time of writing this thesis. The idea was to determine whether *LacZ* expression continued to be directed to appropriate tissues within the head of the embryo, notably including the middle ear and proximal Meckel's cartilage which coincide with *Bapx1*-expressing domains. The results may potentially shed further light not only on the phenotypic effect of loss of *Bapx1* but also, by extrapolation, on the developmental role of *Bapx1*. Such a study also addresses the question of whether *Bapx1* expression is maintained by a self-regulatory loop via binding to the ProxB element. Preliminary results suggest that the latter is not the case since, in the absence of

Bapx1, ProxB-driven *LacZ* continues to be directed to the middle ear/jaw domain (data not shown).

4.3.3.2 Parallels between DistB1- and *Bapx1*-directed expression

The data supporting the designation of DistB1 as a *Bapx1* enhancer is more complex than that for ProxB for a number of reasons. These include the considerably more extensive *LacZ* staining pattern, the inability to detect conservation of the DistB1 element to fish and the fact that the DistB1 transgenic mice only harbour approximately a quarter of the 8kb highly conserved cluster originally identified (see Chapter 3) which may have an important bearing on the observed expression pattern.

Appropriate reporter gene expression is seen in the digits of three independent lines of DistB1-containing transgenic mice which is highly unlikely to be a chance occurrence. Rather, the DistB1 element would appear to be responsible for directing expression to this region. Although the band of *Bapx1* expression seen at E11.5 in fore- and hindlimbs is not recapitulated in the DistB1 lines, *LacZ* staining flanking cartilaginous condensations of the future fore- and hindlimb digits is apparent by E12.5. Despite the early discrepancy, by E14.5, the *Bapx1* expression pattern in cartilaginous condensations of the digits detected by a radioactive *in situ* probe (Tribioli *et al.*, 1997) resembles the *LacZ* staining of the digits of E14.5 DistB1 mice. *Bapx1* and DistB1 expression would appear to be located around the joint regions partitioning the proximal-middle and middle-distal phalanges. The proximal band of expression in the digits both appears first and is the stronger of the two bands for both endogenous *Bapx1* expression and in all DistB1 mice. Notably, this faithfully reflects the chronological order of ossification of the digit bones, there being a one day lag between the ossification of the proximal and distal phalanges of both fore- and hindlimbs (Rugh, 1990).

A further domain of weak expression which is apparent by E15.5 (in line A266.3) appears to correspond to the joint between the proximal phalanges and the metacarpal (forelimb)/metatarsal (hindlimb) bones. This is seemingly reminiscent of a similar region of staining exhibited by $Bapx1^{LacZ}$ mice (Akazawa *et al.*, 2000).

Variation in copy number of the transgene between different DistB1 lines and/or a positional effect exerted by the integration site of the transgene could account for the strongest digit expression being seen in line A266.3. Summarising the above, the principal discrepancy between endogenous *Bapx1* digit expression and that of the reporter gene thus seems to be in terms of temporal regulation. This is potentially explained by the existence of other regulatory elements in close proximity to the DistB1 element under investigation which are necessary to achieve the *bona fide* *Bapx1* pattern. In this respect, the remainder of DistB, or parts thereof, constitute good candidates for additional control elements.

Another region of *LacZ* staining reminiscent of endogenous *Bapx1* expression is that seen in the joints of the long bones within the limb. Although not detected by wholemount *in situ* hybridisation, *Bapx1* expression in limb joints is revealed by *in situ* hybridisation using sectioned embryos (Tribioli *et al.*, 1997). The latter clearly demonstrates *Bapx1* in cartilaginous condensations of the limb at E12.5, for which the pattern closely resembles the *LacZ* staining observed in DistB1 mice at E12.5. In both cases, expression appears to correspond to condensing regions around the humerus and ulna of the forelimb, whilst staining in the hindlimb appears to be located in the vicinity of the tibia and fibula.

As development proceeds, the principal sites of reporter gene expression within the limbs are evidently the joints, as clearly demonstrated by sagittal sections of E15.5 fore- and hindlimbs. These sections also detect, at a lower level, reporter gene expression in the ulna (forelimb) and tibia (hindlimb). Notably, the limbs of DistB1 mice resemble very closely those of $Bapx1^{LacZ}$ mice, the latter exhibiting strong expression in the limb joints with weaker expression in limb bones. Furthermore, the strongest joint expression in $Bapx1^{LacZ}$ mice appears to be the radial-humeral joint of the forelimb, as for DistB1 mice. Thus perhaps the staining observed at early developmental stages along the anterior tip of the forelimb is the precursor of this later joint expression. The difference in temporal regulation of expression with regard to *Bapx1* could be explained, as discussed above, by a requirement for the presence of regulatory sequences additional to the region contained within the DistB1 transgene. It should be noted that reporter gene expression was also detected in the forelimb joint of another line besides A266.3 at

E13.5. This observation further strengthens the evidence that DistB1 drives expression in the limb joints and, by extrapolation, adds sustenance to the argument that DistB1 exerts a regulatory influence on *Bapx1*.

Three independent DistB1 lines display *LacZ* staining in the gut region, a site in which *Bapx1* plays an essential developmental role. Endogenous expression is detectable by E8.5 in lateral plate mesoderm adjacent to the endoderm of the prospective gut (Tribioli *et al.*, 1997). The splanchnic mesoderm staining in $Bapx1^{LacZ}$ mice is also clearly visible during early development (Akazawa *et al.*, 2000). In contrast, gut-specific expression directed by DistB1 did not become apparent until E11.5. At this stage, although staining in the three lines was confined to the posterior stomach, it was not encompassed within exactly the same boundaries as *Bapx1* expression and at no stage was it detected in the spleen, unlike *Bapx1*, which is critical for spleen development. Each of the three DistB1 lines displays staining of the posterior stomach to a varying degree at E11.5 but no line collectively stains the entire posterior stomach. Notably however, staining in line A266 appears to terminate abruptly at the boundary between the posterior stomach and the pyloric sphincter. This is reminiscent of the exclusion of endogenous *Bapx1* from the pyloric sphincter.

Only in line A266.3 does the reporter gene continue to be expressed in the posterior stomach until E14.5, after which it is no longer detected. One of the other two lines exhibits gut expression at E11.5 only, whilst in the other, expression switches to the inner edge of the anterior stomach at E12.5 through E14.5, after which it too is no longer detected. Chance expression in the gut as driven by DistB1 is a possibility, perhaps as a result of this enhancer influencing one of the genes in the *Bapx1* genomic region that is assumed to be ubiquitously expressed. The influence on gene expression exerted by local sequences at the transgene integration site is less likely to account for gut expression in three individual lines (which result from three independent insertional events). It is equally possible that the DistB1 transgene constitutes only part of the genetic circuitry required to direct appropriate spatial and temporal *Bapx1* gut expression. In this respect, DistB1 may be capable of directing reporter gene expression to the appropriate organ (the gut), but acting in isolation is perhaps insufficient to effect the necessary spatial and temporal control.

The latter is envisioned to require the presence of additional regulatory sequences to which as yet uncharacterised upstream and downstream regulatory molecules may bind.

The inability of the DistB1 or ProxB transgenes to drive expression in the spleen suggests that, if these sequences do constitute *Bapx1* regulatory elements, there must be multiple CNSs controlling *Bapx1* expression, one of which that has not herein been identified and is responsible for the spleen expression. Perhaps (but not necessarily), considering the close association of the spleen and the posterior stomach during embryogenesis, the 'spleen element' lies within the 8kb DistB region, adjacent to the proximal part that forms DistB1.

Early detection of *LacZ* expression in the otic vesicle in line A266.3 and later as a patch in the developing ear seen in two separate lines, is again perhaps reflective of the DistB1 element directing expression to the correct organ, yet not in sufficiently specific domains to mimic endogenous *Bapx1* expression (unlike ProxB).

The other tissues in which reporter gene expression is detected are inappropriate with respect to endogenous *Bapx1* expression. The intense, persistent staining of the dorsal root ganglia in the spinal region, in addition to the prominent staining of specific regions of the brain in line A266.3, is perhaps indicative of a positional effect whereby the transgene has integrated nearby sequences responsible for gene expression in the central nervous system. The reason for expression in the spinal region from E12.5, but not always in the dorsal root ganglia (and never the vertebral column) in one other line besides A266.3 is unknown.

Other ectopic regions of expression with regard to endogenous *Bapx1* domains, such as the nose, tongue (data not shown), genitals, heart and lungs are primarily restricted to line A266.3. This again alludes to a positional effect; it would appear that the A266.3 transgene has integrated nearby a particularly strong promoter or enhancer. This question could be addressed by mapping the genomic location of the transgene insertion. An initial approach could be the use of FISH (fluorescence *in situ* hybridisation) for which high resolution techniques such as fibre-FISH have been developed. Interestingly, expression in Rathke's pouch and the tongue has also

been reported for *Nkx3.1* (Tanaka *et al.*, 1999), but whether this has any bearing on the expression pattern of line A266.3 is unknown.

The observed expression pattern is not consistent with that expected of a ubiquitous promoter/enhancer, further strengthening the proposed connection with *Bapx1* as opposed to *RAB28*, *Q9P2L9* or *NM_148894*. It is accepted however that inclusion of the entire 8kb DistB element in a transgenic reporter assay may reveal an expression pattern compatible with one of the aforementioned, seemingly ubiquitously expressed, genes. Such genes typically require only a single, upstream promoter region and, considering the report of a *RAB28* isoform expressed predominantly in the testis (Brauers *et al.*, 1996), it is perhaps significant that the DistB1 element drives *LacZ* expression to such a high level in the genitals. On the other hand, this genital expression is confined to the A266.3 line which renders the link to *RAB28* rather tenuous.

If the 8kb DistB CNS is a genuine *Bapx1* control element, one explanation for the apparent lack of conservation to fish is that it constitutes an enhancer in vertebrate species that have evolved since the common ancestor of mammals and fish. Additional sequence from amphibian and avian species would be extremely useful in addressing such a question. Considering the ancient role of *Bapx1* in the gut, for which conserved genetic elements would perhaps be expected to exist, this invokes the existence of additional 'gut elements' besides a general tendency to direct expression to this organ (as seen for DistB1). The DistB1 transgene also fails to drive expression in the somites and axial skeleton which are critical regions of vertebrate *Bapx1* expression. Investigation by transgenic bioassay of the remaining short, conserved elements contained within the DistB cluster would be predicted to shed considerable light on the question of there being one or more *Bapx1* regulatory elements within the 8kb. Such studies would be further aided by the availability of additional sequence from vertebrate species positioned evolutionarily between the fish and mammalian lineages, such as *Xenopus* and chick. Identification of DistB and/or ProxB sequences in these species could potentially greatly strengthen the link to *Bapx1*, depending on the results of the additional transgenic bioassays proposed to investigate DistB.

4.3.3.3 The potential existence of a complex regulatory network

Overall, there is good evidence to infer a link between the ProxB CNS and regulation of *Bapx1*. Furthermore, the detection of *LacZ* expression in the digits of three, the joints of two and the gut of three DistB1 lines favours the proposal that this enhancer influences *Bapx1* expression, although clearly this partial DistB CNS does not act in isolation. Perhaps it is out of context without its usual proximal elements, some of which may exert inhibitory influences on *Bapx1* to refine the ultimate expression domains and all of which must function combinatorily to ensure rigorous spatial and temporal control of gene expression. The possibility that there exist multiple *Bapx1* regulatory elements is perhaps not surprising given the number of tissues co-opting *Bapx1* expression during evolution since its presumed ancestral role in the visceral mesoderm of the *Drosophila* gut. The results presented here thus point to a highly complex regulatory network whereby multiple sequences must act in concert to ultimately restrict *Bapx1* expression to specific domains of particular organs in a tightly controlled temporal manner.

Chapter 5 The characterisation of molecular boundaries within the gut

5 The characterisation of molecular boundaries within the gut

5.1 Introduction

Defining the roles of the gut-associated mesenchyme and identifying factors required for specialised regional function are central to our understanding of normal gut development. A number of the tissue interactions which direct gut organogenesis have been defined, leading to recognition of the importance of epithelial-mesenchymal signalling. However the molecular nature of the signals responsible for these inductive events is only starting to be elucidated.

Chapters 3 and 4 relate to investigations designed to explore the role of the mesenchyme by exploiting the fact that *Bapx1* is expressed in this tissue. The current chapter details complementary studies into the function of *Bapx1* in patterning events critical to gut development by comparing the gut-specific expression of molecular markers in wildtype mice to those lacking a functional *Bapx1* gene. Analyses of model organisms having undergone genetic manipulation to 'knockout' a particular gene are not only crucial to assigning functions to the gene of interest, but are also instrumental in elucidating putative interacting factors. Furthermore, certain tissue-specific phenotypes associated with the loss of a given gene are often missed in the initial assessment of the mutant, requiring instead a much closer analysis of a particular organ. Indeed in the case of *Bapx1*, for which the generation of null mutants was first reported in 1999 (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999), a role in the murine middle ear has only recently been revealed (Tucker *et al.*, 2003, submitted) and novel manifestations of the mutation within the gut, some of which are reported in the current chapter, are similarly recent findings (manuscript in preparation).

The patterning of the gut into morphologically distinct regions results from the expression of appropriate factors in strict spatial and temporal patterns to assign cells their developmental fates. Previous studies encompassing a range of organisms as evolutionarily diverged as *Drosophila* and mouse have implicated a number of

conserved gene families in various aspects of gut development. These include the *hedgehog* (*hh*) (Bitgood and McMahon, 1995; Roberts *et al.*, 1995; Narita *et al.*, 1998; Murakami *et al.*, 1999), transforming growth factor beta (TGFβ) (Smith and Tabin, 1999; Branford *et al.*, 2000; Smith *et al.*, 2000a) and NK homeobox gene families (Azpiazu and Frasch, 1993; Akazawa *et al.*, 2000; Smith *et al.*, 2000a; Nielsen *et al.*, 2001). Accordingly, a variety of genes expressed within specific domains of the foregut were selected from the aforementioned families in order to analyse their expression pattern in the *Bapx1* mutant gut. It was envisaged that this may reveal the extent to which *Bapx1* is involved in the delineation of, and the importance of maintaining, distinct molecular boundaries in this region during gut development.

5.2 Results

5.2.1 Analysis of *Shh* expression in the *Bapx1* mutant gut

The importance of *Shh* activity to gut morphogenesis, cytodifferentiation and the maintenance of a differentiated state is now well established in vertebrate species (Apelqvist *et al.*, 1997; Ramalho-Santos *et al.*, 2000; Fukuda and Yasugi, 2002; Fukuda *et al.*, 2003; Hebrok, 2003). In both the mouse and chick gut, *Shh* is expressed in the endodermal germ layer prior to gut tube formation (Echelard *et al.*, 1993; Narita *et al.*, 1998). It is later expressed uniformly throughout the gut epithelium (Roberts *et al.*, 1995; Narita *et al.*, 1998) excepting some associated glands and ducts, such as the pancreas, where its downregulation is critical to normal organogenesis (Apelqvist *et al.*, 1997; Hebrok *et al.*, 1998). Within the developing mouse foregut, by E11.5, *Shh* expression defines specific domains such that it is present at a high level in the forestomach epithelium (Bitgood and McMahon, 1995) and absent from the posterior stomach (Kim *et al.*, 2000). Strong expression resumes in the duodenum immediately posterior to the pyloric sphincter which marks the foregut-midgut junction. The *Bapx1* expression pattern is essentially reciprocal to that of *Shh*, being confined to the posterior stomach and spleen, although it is also

undetectable in the pyloric sphincter. It was thus of interest to investigate whether the *Shh* domain was affected by the *Bapx1* mutation.

5.2.1.1 Loss of the posterior stomach boundary

Examination of *Shh* expression was additionally prompted by the fact that in wildtype mice, *Shh* marks a distinct boundary between the uppermost portion of the duodenum and the pyloric sphincter at the posterior end of the stomach. Thus by analysing the *Shh* expression pattern in the absence of functional *Bapx1*, in which case the pyloric sphincter is lost (Akazawa *et al.*, 2000), we could explore the effect of loss of *Bapx1* on molecular boundaries and, concomitantly, on gut morphogenesis.

Wholemount *in situ* hybridisation was performed on dissected guts of wildtype and mutant *Bapx1* littermates at four successive stages of embryonic gut development. Representative results are shown in Figure 5.1. No differences in the wildtype expression pattern were detected between homozygotes and heterozygotes. Endodermal expression of *Shh* within the anterior stomach was sometimes obscured by the trapping of probe in the lumen of the stomach (for example Figure 5.1F). Nevertheless, the anterior expression domain did not appear to be affected by the *Bapx1* mutation, as illustrated by Figure 5.1I. In contrast, the posterior expression domain separating the duodenum from the posterior stomach was clearly defined such that non-ambiguous comparisons between wildtype and mutant guts could be undertaken.

The altered *Shh* expression profile in the absence of *Bapx1* reveals a pronounced morphological transformation whereby expansion of the posterior stomach, together with fusion to the duodenum has occurred (Figure 5.1H-K). Rather than *Shh* expression beginning immediately posterior to the pyloric sphincter, this latter structure is lost in *Bapx1*-/- mice and *Shh*-expressing cells are evident along the caudal edge of the stomach either side of the point at which the newly-positioned duodenum leaves the stomach. This reorganisation of tissue is evident in all *Bapx1* null mutants such that part of the duodenum appears to have become incorporated into the posterior stomach, accounting for the expansion of the latter

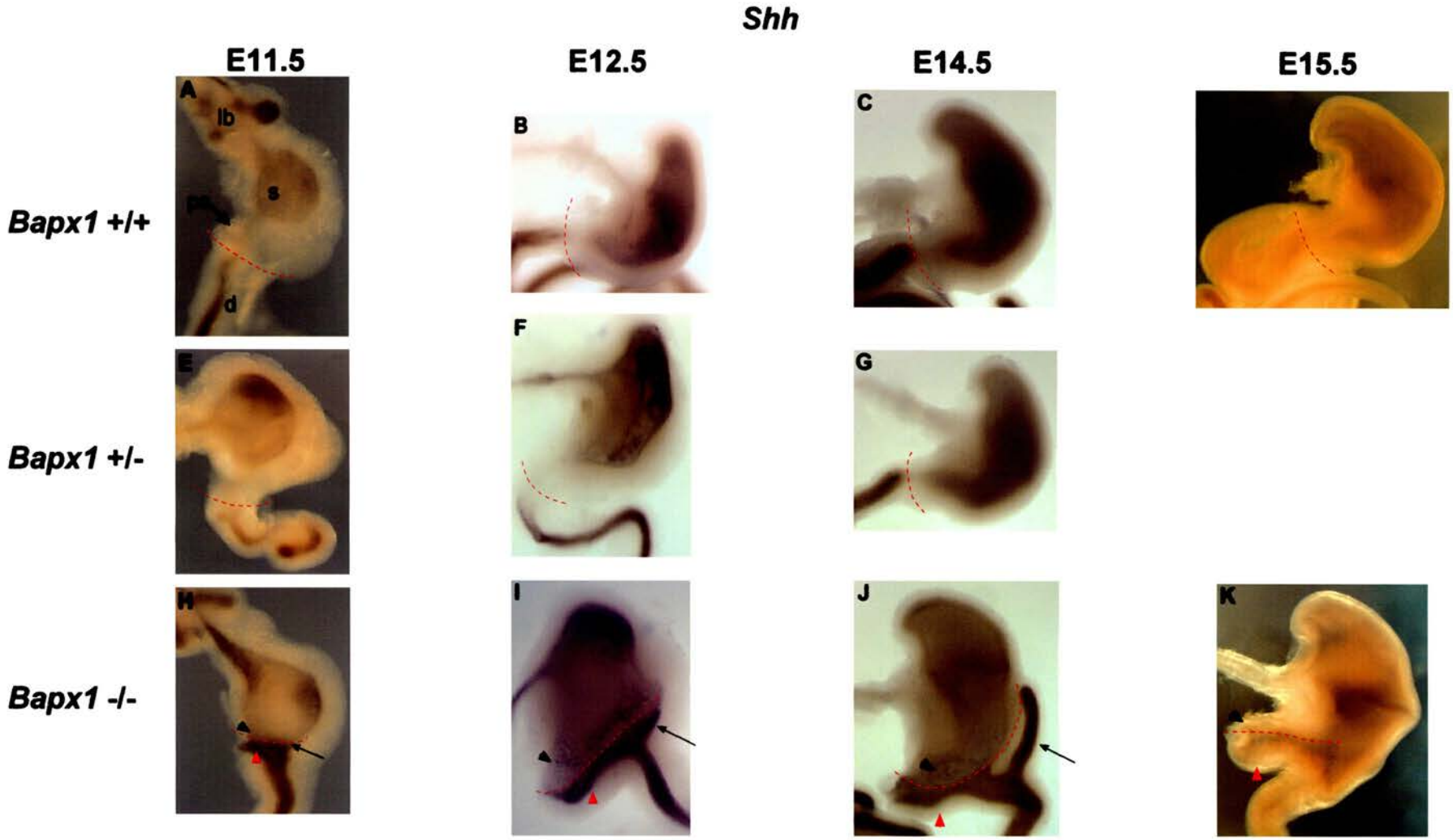


Figure 5.1: *Shh* expression reveals gross morphological changes within the *Bapx1* mutant gut. Detection of *Shh* expression by wholemount *in situ* hybridisation of *Bapx1* +/+ (**A-D**), +/- (**E-G**) and -/- (**H-K**) guts at four different stages of development. *Shh* is expressed in the anterior stomach and duodenum. The spleno-pancreatic region is shown from a left lateral perspective in each case. The red dotted lines indicate the anterior boundary of the duodenum. In mutant (-/-) *Bapx1* guts, expansion of the posterior stomach is observed (black arrowhead) (**H-K**). In contrast to wildtype guts, *Bapx1* mutants also display abnormal folding of the duodenum (red arrowheads), together with fusion to the posterior stomach. At E11.5 (**H**), E12.5 (**I**) and E14.5 (**J**), branch structures (indicated by black arrows) are visible emanating from the duodenum in *Bapx1* mutants. d, duodenum; lb, lungbuds; ps, pyloric sphincter; s, stomach.

(compare *Bapx1* wildtype (+/+ and +/-) guts to mutant (-/-) guts; Figure 5.1). Nevertheless, the fused small intestine appears to have retained its cellular identity.

The sharp distinction between the posterior stomach and duodenum seen in wildtype mice has clearly been significantly disrupted in the mutant gut. Not only has tissue been repositioned, but the molecular boundary between *Shh*-expressing and non-expressing cells in the mutant also appears to be somewhat blurred. Figure 5.1I in particular demonstrates 'speckling' of *Shh*-expressing cells anterior to the duodenal fusion but contained within the enlarged posterior stomach of the mutant. This raises the intriguing question as to whether any posterior stomach tissue has also been re-specified to adopt a duodenal fate. To address this prospect, E12.5 and E14.5 wholemount guts were sectioned for closer analysis. However, owing to the close proximity of the *Shh*-expressing cells and the posterior gut lumen in the mutant, it was unclear whether trapping of the probe accounted for the resultant staining pattern (data not shown). Nonetheless, further insight into the posterior boundary loss was facilitated by a close inspection of *Bapx1* mutant guts stained with the intestinal enzyme alkaline phosphatase. At E17.5, alkaline phosphatase appears to have extended its domain anteriorly into the posterior stomach of the *Bapx1* mutant gut (Figure 5.2; Hill group, unpublished data). As revealed by the altered *Shh* expression profile, the distinct boundary separating the small intestine from the posterior stomach is absent from *Bapx1*-/- guts and alkaline phosphatase is detected in the posterior stomach (Figure 5.2C, D). This provides a further indication that an alteration to molecular domains accompanies the morphological expansion of the mutant stomach. Having lost a well-defined morphological and molecular boundary, cell mixing appears to occur in the region of the expanded posterior stomach (Figure 5.2C), providing a plausible explanation for the observed *Shh* 'speckling' pattern without the need to invoke re-specification of cellular identity.

5.2.1.2 An abnormal gut branching phenotype

A further manifestation of the loss of *Bapx1* revealed by *Shh* expression is an aberrant branching of the duodenum that is apparent by E12.5 and is observed in

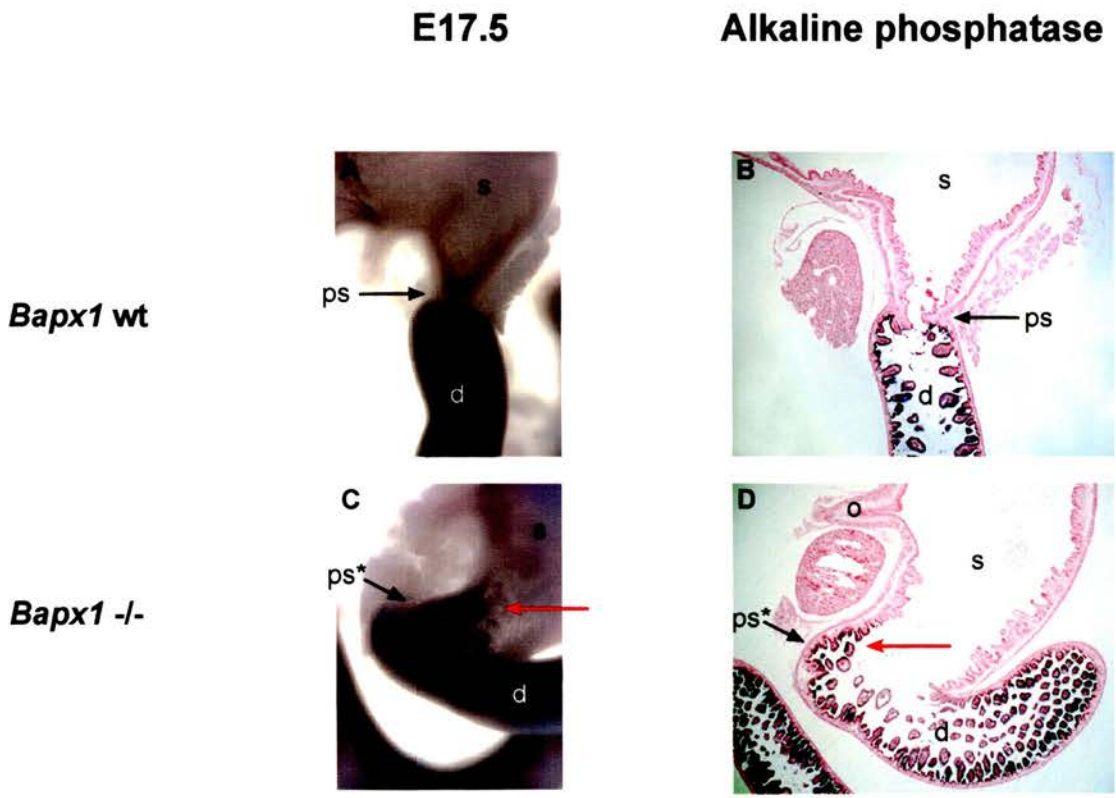


Figure 5.2: Alkaline phosphatase staining reveals the loss of a distinct boundary between the posterior stomach and duodenum within the *Bapx1* mutant gut at E17.5. *Bapx1* wildtype (wt) (**A** and sagittal section in **B**) and mutant (**C** and sagittal section in **D**) E17.5 guts stained with the intestinal enzyme alkaline phosphatase. A sharp boundary is evident at the foregut-midgut junction, marked by the pyloric sphincter (ps), in wt guts (**A**, **B**). In the absence of *Bapx1* (**C**, **D**), the alkaline phosphatase stain spreads rostrally into the posterior stomach and indicates that the foregut-midgut junction has become a region of cell mixing (red arrow). The position at which the pyloric sphincter would be expected in the *Bapx1* -/- gut is indicated (*). d, duodenum; o, oesophagus; ps, pyloric sphincter; s, stomach. Images derive from the group of Dr. Robert Hill, MRC Human Genetics Unit, Edinburgh; unpublished data.

approximately 50% of homozygous null mutants (Figure 5.1I, J). At E12.5, the branch appears to be a continuous extension of the duodenal region fused to the posterior stomach (Figure 5.1I). However during subsequent development of the region, the branch extends in an anterior direction along the greater curvature of the stomach wall and by E14.5 endodermal *Shh* expression within the duodenal branch demonstrates that it exists as a separate entity to the posterior stomach (Figure 5.1J). The distinct physical separation of *Shh*-expressing cells in the mutant branch from the adjacent tissue of the posterior stomach is further illustrated by still images of both wildtype (Figure 5.3A) and mutant (Figure 5.3B) E14.5 wholemount guts processed for Optical Projection Tomography (OPT) analysis.

Transverse sections through the branching structure of the E14.5 mutant gut (Figure 5.3C-E) demonstrate that, moving in an anterior to posterior direction, the majority of the duodenal branch constitutes a separate structure and retains an individual lumen (Figure 5.3C). At a more posterior level, the branch fuses with the posterior stomach (Figure 5.3D) such that a common gut lumen exists (Figure 5.3E). The origin of the anomalous outgrowth lies in close proximity to the duodenal region fused to the posterior stomach (Figure 5.3B) and indeed the branch itself would appear to be attached to the adjacent tissue of the dorsal stomach. Perhaps this attachment is required to support and/or direct the branch extension. The fate of this anomalous growth after E14.5 has not yet been determined due to the limited availability of mutant tissue and to the incomplete penetrance of the branching phenotype.

The apposition, in *Bapx1* mutant guts, of *Shh*-expressing endodermal tissue and *Bapx1*-/- mesenchymal tissue of the posterior stomach suggests that the abnormal branching phenotype constitutes an endodermal response to a mesenchymal deficiency. The absence of *Bapx1* in the mesenchyme is presumably responsible for an altered response of downstream signalling molecules to morphogenetic cues. The result, in addition to asplenia, is a loss of the posterior stomach boundary and the potential for an abnormal duodenal branch to develop alongside the posterior stomach.

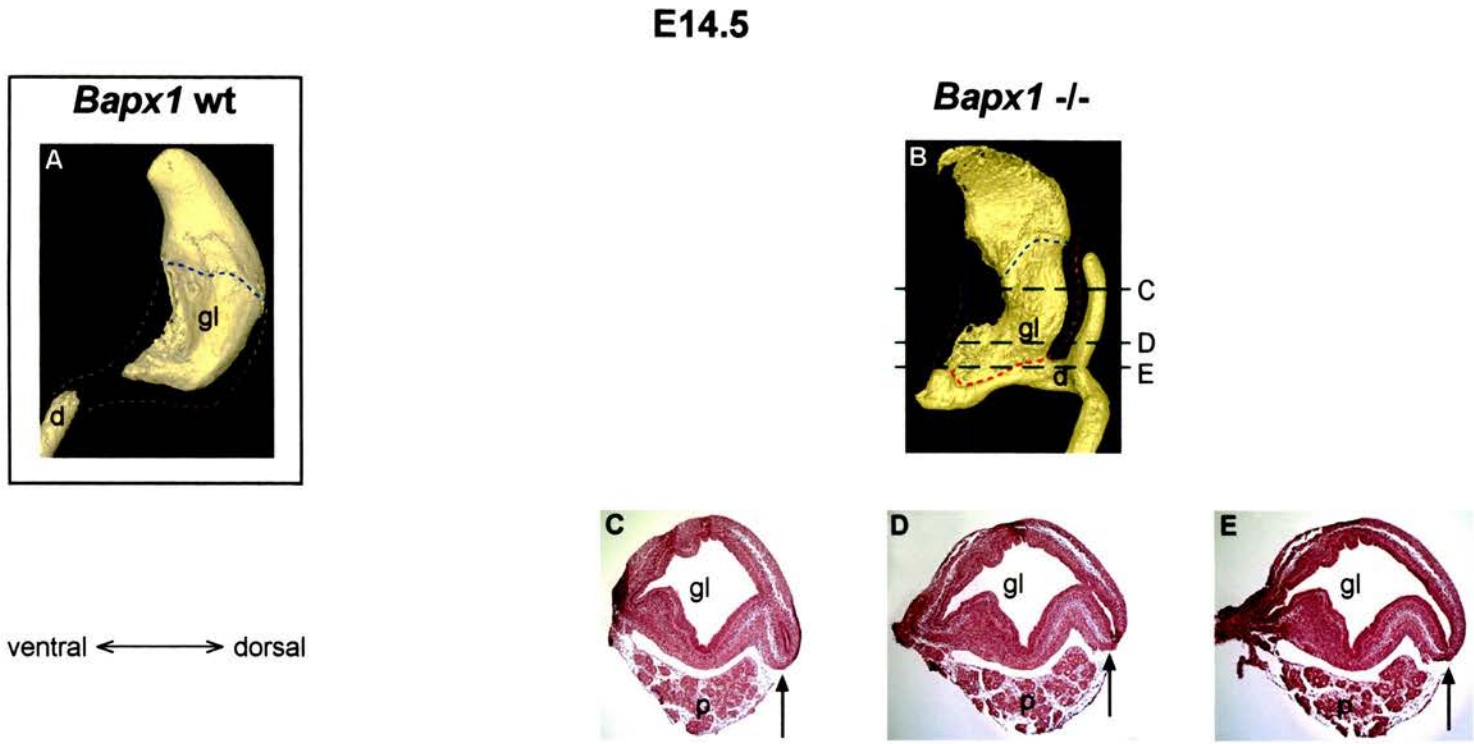


Figure 5.3: Further characterisation of the *Bapx1* mutant gut. *Bapx1* wildtype (wt) (**A**) and mutant (**B**) E14.5 guts hybridised with a *Shh* probe (see Figure 5.1) were subjected to OPT analysis which served to underscore the morphological transformation of the mutant gut. *Shh* expression in the anterior stomach appears above the blue dotted lines. Pooling of probe in the gut lumen (gl) renders this tissue visible. The posterior stomach, from which *Shh* is excluded, is outlined by yellow dotted lines. Transverse sections (**C-E**) taken through the abnormal branching structure of the E14.5 *Bapx1* -/- gut at the positions indicated in **B** demonstrate that the branch (denoted by a black arrow) lying alongside the dorsal stomach is a distinct structure with a separate lumen (**C**). At a more posterior level, the branch has clearly fused with the posterior stomach to form a continuum whereby a common gut lumen exists (**E**). d, duodenum; gl, gut lumen; p, pancreas.

5.2.2 Investigation of *Bapx1* and *Nkx2.5* in the *Shh* mutant gut

Profound disruption of molecular boundaries and gross morphological changes were revealed in the *Bapx1* mutant gut by the altered *Shh* expression pattern. Nevertheless, *Shh* expression was not lost in the absence of *Bapx1*, persisting in both the anterior stomach and duodenal cells albeit in their novel location in the case of the latter. This suggests that *Shh* is not downstream of *Bapx1* in a signalling pathway and is not the molecule responsible for effecting the morphological transformation. A number of lines of evidence drawn from studies of evolutionary distant organisms suggest that *Shh* is upstream of *Bapx1*. Firstly, in *Drosophila*, epithelial *hedgehog* is required for *bagpipe* expression in the gut mesoderm (Hoch and Pankratz, 1996) and various investigations suggest that features of the *Drosophila* and vertebrate *hedgehog* signalling pathways are conserved. For example, ectopic expression of zebrafish *Shh* is able to phenocopy the effects of *Drosophila hedgehog* (Krauss *et al.*, 1993; Ingham and Fietz, 1995). Secondly, a number of studies relating to gut development have demonstrated that epithelial *Shh* signals to the adjacent mesenchyme to induce differentiation (Roberts *et al.*, 1995; Apelqvist *et al.*, 1997; Ramalho-Santos *et al.*, 2000). Furthermore, as mentioned previously, the expression pattern of *Shh* in the spleno-pancreatic region is essentially reciprocal to that of *Bapx1*. It was therefore of interest to determine the effect of an absence of *Shh* upon *Bapx1* expression domains in the gut.

To further explore the nature of the relationship between *Shh* and *Bapx1*, and to investigate whether comparable molecular boundaries were disrupted in the absence of *Shh*, the *Bapx1* expression pattern in the *Shh* mutant gut was determined by wholemount *in situ* hybridisation. In the wildtype gut, *Nkx2.5* (in addition to spleen expression) is spatially restricted to the mesoderm of the pyloric sphincter therefore provides a convenient marker for a domain located directly between, but not overlapping, *Shh* expression in the duodenum and *Bapx1* expression in the posterior stomach. *Nkx2.5* was therefore included in the present analysis as an additional means of assessing the effect of *Shh* absence on molecular boundaries within the foregut. A panel illustrating the endogenous *Bapx1* expression pattern in

wildtype dissected guts between the ages of E10.5 and E13.5 is included in Appendix 2.

5.2.2.1 The *Bapx1* and *Nkx2.5* expression domains remain constant

Three successive stages of foregut development for which *Bapx1* and *Nkx2.5* domains are clearly defined were included in the *in situ* hybridisation analysis. No distinction is made between wildtype homozygotes and *Shh* heterozygotes as no difference in the expression pattern was observed. Comparison of wildtype guts to those lacking *Shh* reveals no obvious alteration to either the *Bapx1* or the *Nkx2.5* expression domains at any of the stages examined (Figure 5.4). Lateral views of wildtype and mutant guts (Figure 5.4A-L) demonstrate *Bapx1* and *Nkx2.5* in the posterior stomach and pyloric sphincter respectively. From a dorsal perspective, expression of the two genes is demonstrated in the spleen of both wildtype and *Shh* mutants (Figure 5.4C'-F', I'-L'). (The dark staining observed for *Nkx2.5* in Figure 5.4I, I', K, K', L, L' reflects trapping of the probe in the lumen of the stomach). These results indicate that the *Bapx1* and *Nkx2.5* genes reside in a separate molecular pathway to *Shh* as neither exhibit altered expression patterns in its absence.

5.2.2.2 An alteration to the morphology of the spleen and the pancreas

Interestingly, a hitherto unreported alteration to the morphological appearance of the pancreas and the spleen is manifest in the *Shh* mutant gut at E13.5 (Figure 5.4F', L'). The anterior portion of the spleen appears to have acquired a more bulbous appearance whilst the dorsal pancreas has adopted a flatter, more elongated shape as compared to its wildtype counterparts. The significance of these morphological changes is unknown, however they suggest that a deficiency of *Shh* may exert an early effect on gut morphogenesis that is independent of the signalling molecule(s) directly responsible for induction of *Bapx1* and *Nkx2.5* expression.

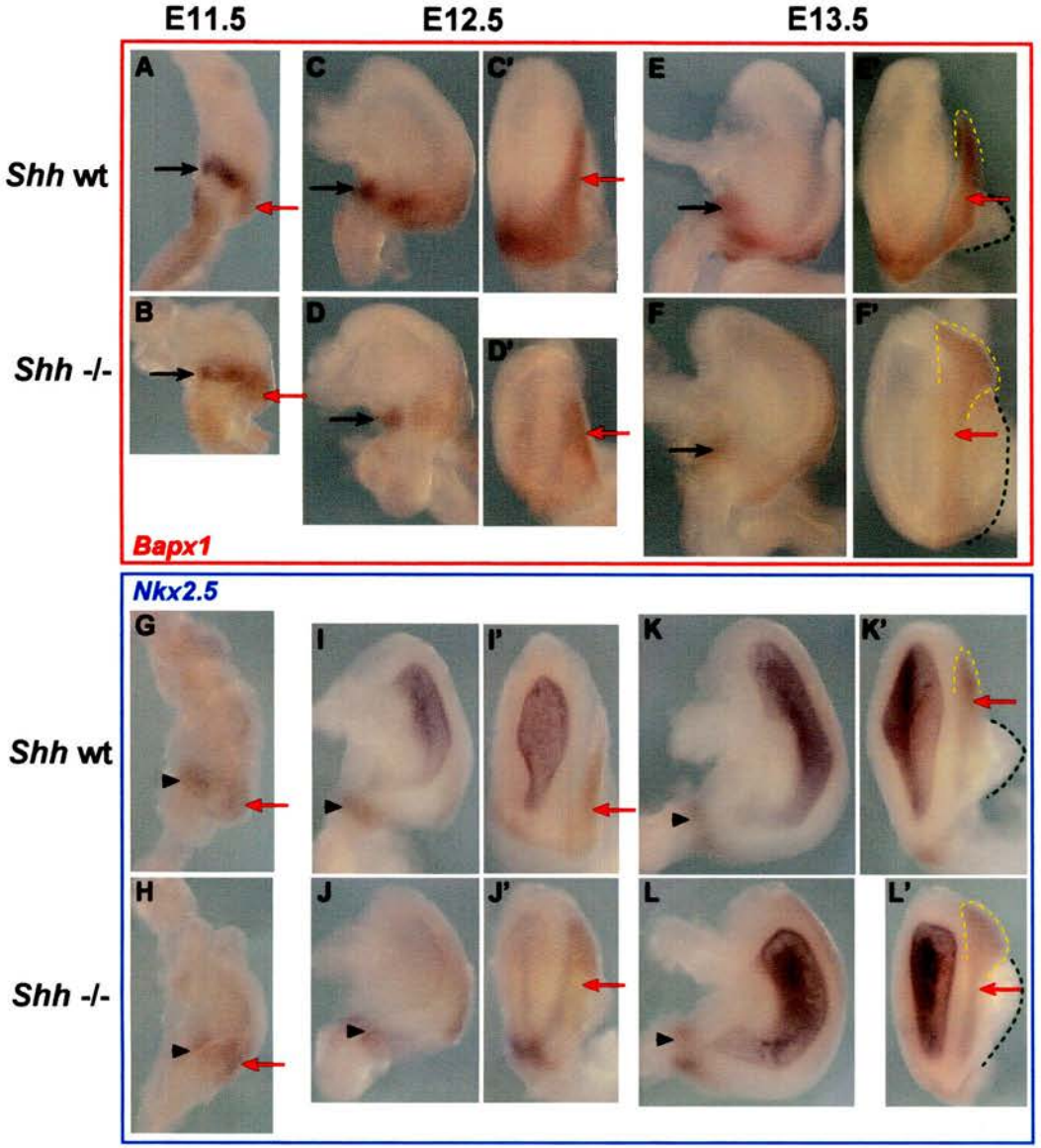


Figure 5.4: No change in *Bapx1* or *Nkx2.5* expression is detected in the *Shh* mutant gut. Wholemount *in situ* hybridisation of E11.5 (**A**, **B**, **G**, **H**), E12.5 (**C**, **C'**, **D**, **D'**, **I**, **I'**, **J**, **J'**) and E13.5 (**E**, **E'**, **F**, **F'**, **K**, **K'**, **L**, **L'**) guts with either a *Bapx1* (**A-F**) or *Nkx2.5* (**G-L**) RNA *in situ* probe. The spleno-pancreatic region is shown from either a left lateral (**A-L**) or dorsal (**C'-F'**, **I'-L'**) perspective. In wildtype (*Shh* wt) mice, *Bapx1* is expressed in the posterior stomach (black arrow) (**A**, **C**, **E**) and spleen (red arrow) (**A**, **C'**, **E'**). These domains of expression do not appear to be altered in the *Shh* -/- guts (**B**, **D**, **D'**, **F**, **F'**). The wildtype expression domains of *Nkx2.5* within the foregut are the pyloric sphincter (black arrowhead) (**G**, **I**, **K**) and the spleen (red arrow) (**G**, **I'**, **K'**). Expression within these tissues is maintained in the *Shh* -/- guts (**H**, **J**, **J'**, **L**, **L'**). Trapping of the *Nkx2.5* probe in the gut lumen accounts for the purple stain seen in **I**, **I'**, **K**, **K'**, **L**, **L'**. The splenic expression of *Bapx1* and *Nkx2.5* reveals an alteration to the morphology of both the spleen and the pancreas in *Shh* mutants that is evident by E13.5 (compare **E'** to **F'** and **K'** to **L'**). The anterior end of the spleen has acquired a more rounded appearance (yellow dotted line), whilst the pancreas has a more elongated shape (green dotted line) in the *Shh* -/- guts.

5.2.3 Investigation of the *Bmps*

The *Bone morphogenetic proteins* (*Bmps*) belong to the TGFβ (Transforming Growth Factor beta) superfamily of signalling molecules and are vertebrate relatives of *decapentaplegic* (*dpp*) in *Drosophila* (Bitgood and McMahon, 1995). Members of the *Bmp* family are often expressed in tissues adjacent to those expressing *hedgehog* genes which reflects an ancient epistatic relationship between these classes of molecules (Bitgood and McMahon, 1995). In the developing chicken gut, epithelial *Shh* is involved in the differentiation of the proventriculus and the gizzard (Narita *et al.*, 1998; Fukuda *et al.*, 2003). It exerts its effect often through the induction of *Bmps* in the mesenchyme, amongst which *Bmp2* (Narita *et al.*, 2000) and *Bmp4* (Nielsen *et al.*, 2001) have been assigned important roles. Notably, SHH can induce expression of *Bmp4* in the midgut (Roberts *et al.*, 1998) and in the hindgut (Roberts *et al.*, 1995) yet ectopic SHH production in the gizzard is unable to elicit *Bmp4* expression (Roberts *et al.*, 1998). The absence of BMP4 from the mesoderm of the developing gizzard, but its presence in the proventriculus, has been suggested to depend upon inhibition by *Bapx1* which is expressed in the gizzard but not in the proventriculus (Nielsen *et al.*, 2001). Owing to the conserved regions of *Bapx1* foregut expression between the avian and mammalian lineages (that is, presence in the posterior stomach but absence from the anterior domain) it was thus reasonable to expect upregulation of *Bmp4* in the posterior stomach of the *Bapx1* null mutant if indeed *Bapx1* does constitute an inhibitor of *Bmp4*. Further support for a relationship between these two molecules stems from observations made during the molecular analysis of the axial skeleton of *Bapx1* mutant mice (Tribioli and Lufkin, 1999). The *Bmp4* expression domain, normally strictly confined to the perichondrial region, was more widespread and homogeneous in the absence of *Bapx1*, again raising the possibility that the latter may regulate *Bmp4*.

Taken together, the reasons above provided the rationale for investigating *Bmp* expression in the *Bapx1* mutant, the idea being to explore the role of *Bapx1* in gut development and simultaneously search for additional factors responsible for the specification of molecular boundaries within this region. An additional premise for the investigation of *Bmp* expression in the *Bapx1* mutant gut stemmed from

observations made during early embryogenesis relating to the formation of the SMP. *Bapx1* was asymmetrically expressed in the left LPM by E10.5 and was implicated in regulating the leftward growth of the spleno-pancreatic mesenchyme (Hecksher-Sørensen *et al.*, 2003, submitted). *Bmp11* was similarly found to be restricted to the left side of the spleno-pancreatic region by E10.5, suggesting that this family of growth factors may play important roles in the asymmetric growth and development of the foregut. The possibility arose that misregulation of the *Bmps* may be a factor in the abnormal growth detected in the posterior stomach region of the *Bapx1* null mutant.

The expression patterns of four *Bmp* probes were compared between *Bapx1* wildtype and mutant littermates by wholemount *in situ* hybridisation. Transcripts corresponding to *Bmp2* and *Bmp7* were detected in appropriate places as previously reported (Lyons *et al.*, 1990; 1995), indicating that the probes were functional, however neither were detected in the E10.5 or E12.5 foregut, either wholemount or sectioned (data not shown). Attention thus focussed on *Bmp4* and *Bmp11* in light of the implied interaction between *Bapx1* and *Bmp4* in the chick, and the asymmetric expression of *Bmp11* detected in the spleno-pancreatic region. Regarding the latter, the expression level was very low (Hecksher-Sørensen, 2001) which may explain why I was unable to identify *Bmp11* transcripts in the foregut (data not shown) and consequently was not able to ascertain the effect of the *Bapx1* null mutation on expression of *Bmp11* in the spleno-pancreatic mesenchyme. Nevertheless, if this molecule were to play a prominent role in the establishment of molecular boundaries within the foregut, one would perhaps expect it to be expressed at an appreciable level before E12.5, at which point the morphological transformation of the *Bapx1* mutant foregut is clearly apparent.

Based on studies of the chick gizzard, an upregulation of *Bmp4* in the posterior stomach of mouse was expected if the role of this molecule in gut development has been conserved between mammals and birds. Expression of *Bmp4* is detected in the limb buds and LPM at E10.5 (Figure 5.5A, B), confirming that the probe is functional. No expression is detected however in either wildtype or mutant dissected guts at E10.5, both before and after sectioning (Figure 5.5C-I). Confocal

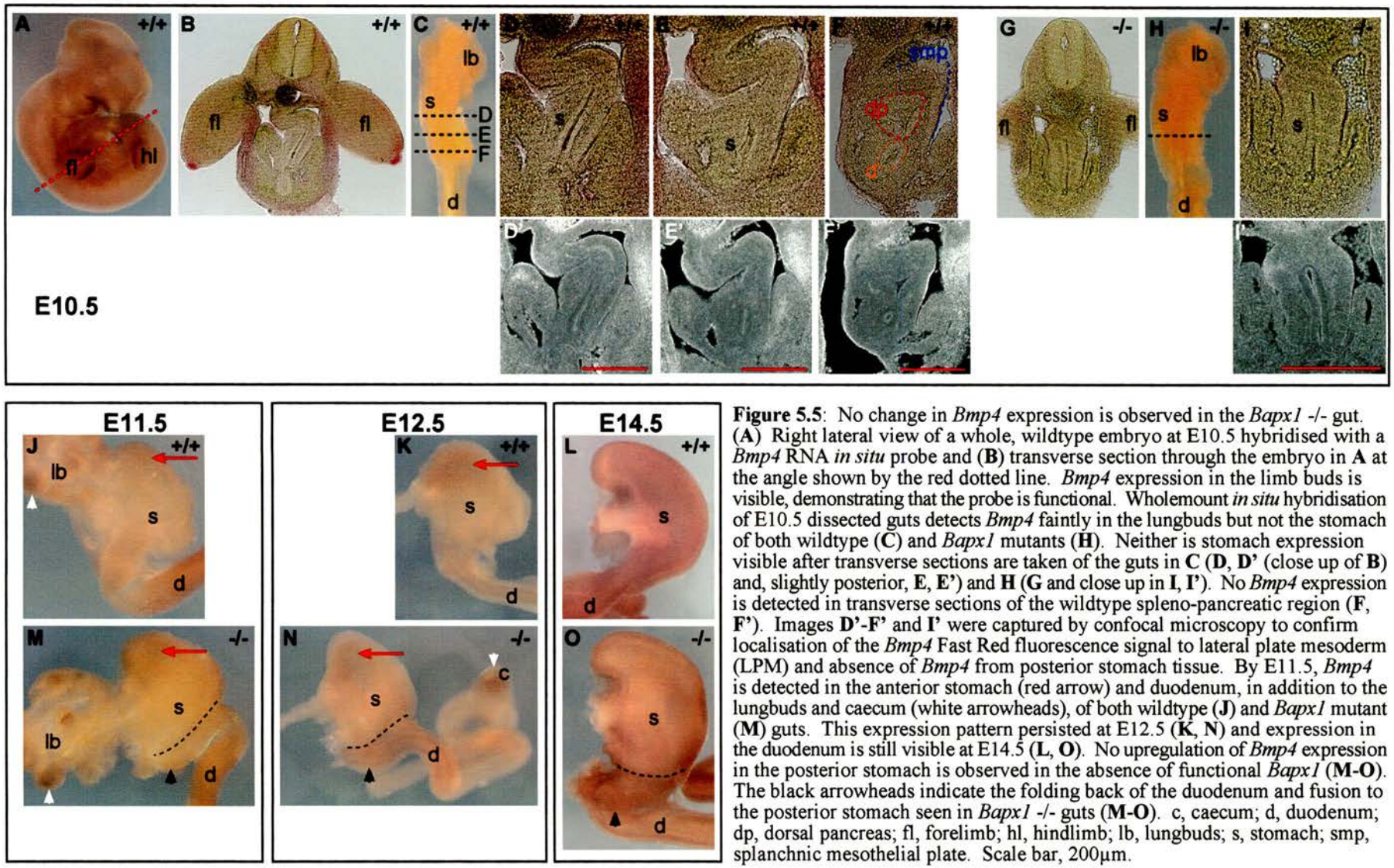


Figure 5.5: No change in *Bmp4* expression is observed in the *Bapx1* -/- gut. **(A)** Right lateral view of a whole, wildtype embryo at E10.5 hybridised with a *Bmp4* RNA *in situ* probe and **(B)** transverse section through the embryo in **A** at the angle shown by the red dotted line. *Bmp4* expression in the limb buds is visible, demonstrating that the probe is functional. Wholemount *in situ* hybridisation of E10.5 dissected guts detects *Bmp4* faintly in the lungbuds but not the stomach of both wildtype (**C**) and *Bapx1* mutants (**H**). Neither is stomach expression visible after transverse sections are taken of the guts in **C** (**D**, **D'** (close up of **B**) and, slightly posterior, **E**, **E'**) and **H** (**G** and close up in **I**, **I'**). No *Bmp4* expression is detected in transverse sections of the wildtype spleno-pancreatic region (**F**, **F'**). Images **D'-F'** and **I'** were captured by confocal microscopy to confirm localisation of the *Bmp4* Fast Red fluorescence signal to lateral plate mesoderm (LPM) and absence of *Bmp4* from posterior stomach tissue. By E11.5, *Bmp4* is detected in the anterior stomach (red arrow) and duodenum, in addition to the lungbuds and caecum (white arrowheads), of both wildtype (**J**) and *Bapx1* mutant (**M**) guts. This expression pattern persisted at E12.5 (**K**, **N**) and expression in the duodenum is still visible at E14.5 (**L**, **O**). No upregulation of *Bmp4* expression in the posterior stomach is observed in the absence of functional *Bapx1* (**M-O**). The black arrowheads indicate the folding back of the duodenum and fusion to the posterior stomach seen in *Bapx1* -/- guts (**M-O**). c, caecum; d, duodenum; dp, dorsal pancreas; fl, forelimb; hl, hindlimb; lb, lungbuds; s, stomach; smp, splanchnic mesothelial plate. Scale bar, 200μm.

microscopy confirms localisation of the Fast Red *Bmp4* signal to LPM (Figure 5.5D'-F') but absence of expression in the posterior stomach (Figure 5.5D', E') or spleno-pancreatic region (Figure 5.5F') of wildtype and *Bapx1* mutant (Figure 5.5I') guts. By E11.5, *Bmp4* expression is visible in the anterior stomach and the duodenum of both wildtype and mutant guts (Figure 5.5J, K, M, N), in addition to the lung buds, as expected from previous reports (Bellusci *et al.*, 1996). Expression of *Bmp4* is also noted in the caecum of wildtype and mutant (Figure 5.5N) guts which provides a further control for probe functionality. The transcript distribution of *Bmp4* in the mouse gut reflects that observed in the comparable regions of the chick gut (Bitgood and McMahon, 1995; Roberts *et al.*, 1998). Contrary to expectation however, close analysis of the gut region does not expose an upregulation of *Bmp4* expression in the posterior stomach in the absence of functional *Bapx1* (Figure 5.5M-O). Instead, *Bmp4* transcript distribution is reminiscent of the *Shh* expression reported above (albeit not as striking as in the case of endodermal *Shh*) in that *Bmp4* is also maintained in the portion of the duodenum fused to the posterior stomach. The inability to detect *Bmp4* expression anterior to the fused region, however, suggests that perhaps *Bapx1* is not directly responsible for *Bmp4* inhibition in the mammalian posterior stomach. Alternatively, there may be redundancy amongst signalling molecules such that another factor can compensate for the absence of *Bapx1*.

5.2.4 Additional genes whose expression patterns demarcate clear boundaries within the developing foregut

5.2.4.1 *Six2* and *Barx1*

Barx1 is a homeobox gene which marks the future stomach region of the primitive gut at E9.5 and is present in the mesenchymal wall of the developing stomach until E16.5 (Tissier-Seta *et al.*, 1995). The homeobox-containing transciption factor *Six2* (Oliver *et al.*, 1995) is expressed in the mesoderm of the posterior stomach and, as in the case of *Barx1*, the expression pattern is conserved between mouse and chick (Smith *et al.*, 2000b). Owing to the clear distinction between the posterior stomach

and duodenum elicited by their wildtype expression patterns, these two genes were selected for analysis of the disrupted boundary in the *Bapx1* mutant. Dissected guts at E12.5 and E14.5 stages of development were hybridised with *Six2* and *Barx1* RNA *in situ* probes. Homozygous and heterozygous wildtype expression patterns are identical for both *Barx1* (Figure 5.6A, B, D, E) and *Six2* (Figure 5.6G, H, J, K); in each case there is a conspicuous boundary at the foregut-midgut junction between expressing and non-expressing cells. Analysis of E12.5 and E14.5 mutant guts reveals no change in regional expression domains despite the morphological transformation of the *Bapx1* mutant gut; neither *Barx1* (Figure 5.6C, F) nor *Six2* (Figure 5.6I, L) stray across the posterior stomach-duodenal boundary. The gross alteration in tissue morphology is likely to account for the seemingly more diffuse *Six2* expression pattern in the absence of *Bapx1* although notably, it remains confined to the posterior stomach (Figure 5.6I, L).

The data presented above suggests that *Barx1* and *Six2* are not responsible for the reorganisation of gut tissue manifest in the *Bapx1* null mutant. Furthermore, the constancy of the *Barx1* and *Six2* expression domains does not thus provide additional information regarding the relationship between the factors responsible for specification of molecular boundaries within the foregut and their effect on tissue morphology. As molecular gut markers however, *Barx1* and *Six2* are informative in that, whilst the distinct boundary between the posterior stomach and duodenum is lost, no alteration to their mesenchymal expression domains within the foregut is apparent. Hence loss of the posterior stomach boundary in the *Bapx1* mutant gut does not appear to effect mesenchymal re-specification.

5.2.4.2 The *Wnt* genes

The *Wnt* genes encode a large family of secreted signalling molecules that, amongst a variety of functions within the embryo, have been proposed to play a role in specifying the gastrointestinal tract (Wells and Melton, 1999). Studies in the chick provided a preliminary indication that *Wnt5a* may be affected by the *Bapx1* mutation in a similar manner to *Bmp4* (Nielsen *et al.*, 2001). Ectopic expression of *Bapx1* in

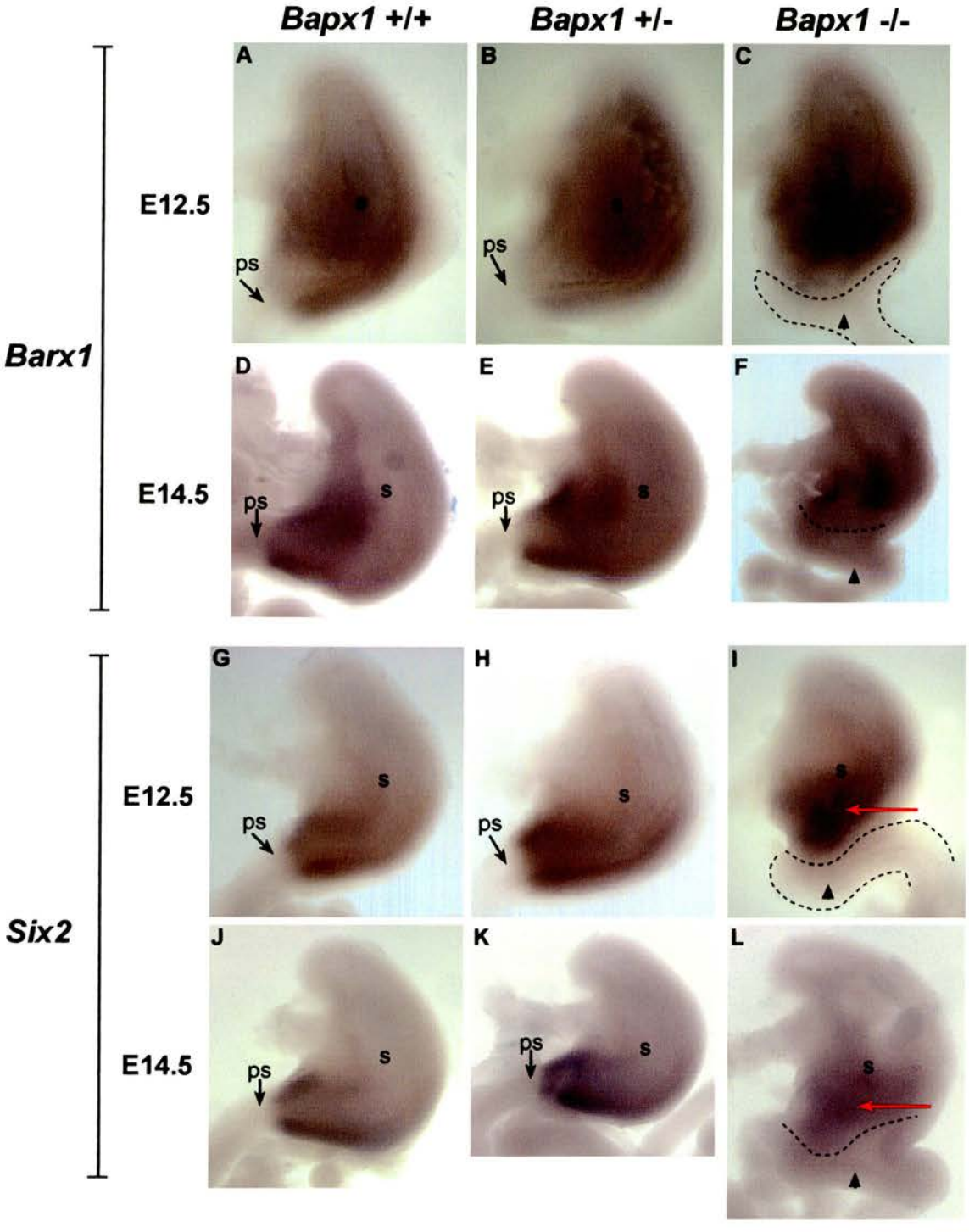


Figure 5.6: No change is observed to the molecular domains of *Barx1* and *Six2* expression within the foregut of *Bapx1* -/- mice. Wholemount *in situ* hybridisation of dissected guts (shown from a left lateral perspective) with *Barx1* (**A-F**) and *Six2* (**G-L**) at E12.5 (**A-C**, **G-I**) and at E14.5 (**D-F**, **J-L**). *Barx1* is expressed in the wall of the developing stomach of wildtype *Bapx1* guts and a distinct molecular boundary is evident at the foregut-midgut junction, marked by the pyloric sphincter (ps) (**A**, **B**, **D**, **E**). The pyloric sphincter is absent in *Bapx1* -/- mice however *Barx1* continues to be expressed in the stomach and excluded from the duodenum (**C**, **F**). Black arrowheads mark the folding back of the duodenum and fusion to the posterior stomach in *Bapx1* -/- guts (**C**, **F**, **I**, **L**). In wildtype guts, *Six2* transcripts localise to the posterior stomach with the strongest expression being observed immediately anterior to the pyloric sphincter (**G**, **H**, **J**, **K**). Although *Six2* expression appears to be more diffuse in the *Bapx1* -/- guts, transcripts remain confined to the posterior stomach (red arrows) and absent from the duodenum (**I**, **L**). ps, pyloric sphincter; s, stomach.

the proventriculus inhibited the normal proventricular expression of *Wnt5a* in addition to *Bmp4*. Based on these results, it was hypothesised that *Bapx1* was a regulator of *Wnt5a* therefore an alteration to the expression pattern of the latter may be observed in the *Bapx1* mutant gut. Accordingly, *in situ* hybridisation was performed on E12.5 mouse guts using a *Wnt5a in situ* probe. Consistent with the wildtype expression pattern in the chick reported by Nielsen *et al.*, *Wnt5a* transcripts were detected in the anterior stomach and the midgut of wildtype mouse guts (data not shown). No appreciable difference was evident however in the *Bapx1* mutant guts and, as in the case of *Bmp4*, no expression was detected in the posterior stomach (data not shown). These results suggest that *Bapx1* is not directly responsible for the exclusion of *Wnt5a* from the posterior stomach. By extrapolation, *Wnt5a* is unlikely to be the effector of the morphological transformation observed in the *Bapx1* mutant gut.

5.3 Conclusions and Discussion

5.3.1 Gross morphological changes and loss of a distinct boundary in the *Bapx1* mutant gut

Molecular analysis of the skeletal phenotype of *Bapx1* mutant mice revealed that *Shh* expression was unaffected in the *Bapx1* null background (Tribioli and Lufkin, 1999). Similarly, in the present investigation of the gut-specific phenotype, the *Shh* domain *per se* is not affected by loss of *Bapx1*, that is *Shh* remains detectable in the anterior stomach and duodenum. However, examination of *Shh* expression in the *Bapx1* mutant gut *does* reveal gross alterations in tissue morphology not previously appreciated. The duodenum, no longer separated from the posterior stomach by the pyloric sphincter due to absence of the latter in *Bapx1*-/- mice, has fused to the posterior stomach, accounting for the apparent expansion of the posterior stomach in the *Bapx1* null background. Furthermore, aberrant branching of the duodenum which extends from its origin in an anterior direction, adjacent to the stomach, is observed in approximately 50% of null mutants. Thus the expression profile of *Shh*

serves to highlight the importance of *Bapx1* to normal patterning of the gut by revealing the extent of the morphological transformation of gut tissue in the absence of *Bapx1*. The curious appearance of the *Bapx1* mutant gut raises fascinating questions as to the molecular signals responsible for this aberrant growth and the reasons for its appearance in some, but not all, *Bapx1* mutants.

Seemingly punctate *Shh* expression is evident immediately anterior to the novel posterior stomach-duodenum fusion. Taken alone, this may indicate re-specification of cellular identity within the posterior stomach such that the posterior-most cells adopt a duodenal fate. However, in guts stained to detect the intestinal enzyme alkaline phosphatase, the distinct boundary between non-expressing foregut-derived tissue and expressing duodenal tissue is also evidently lost, but is replaced by a gradual fading of stain. Thus taken together, the 'speckling' *Shh* pattern and gradual fading of alkaline phosphatase stain are more consistent with cell mixing due to the loss of a distinct boundary. In wildtype guts, the pyloric sphincter forms a specialised tissue at the junction between the foregut and midgut and *Nkx2.5* exhibits a temporally and spatially restricted expression pattern within the pyloric sphincter (Smith and Tabin, 1999). It is thus plausible that in the absence of *Bapx1*, in which case *Nkx2.5* and the pyloric sphincter are lost, the disappearance of both a morphological and molecular boundary are conducive to cell mixing by removing physical and potentially inhibitory molecular barriers. Interestingly, recent data from our lab suggests that the removal of a physical barrier *per se* is not sufficient to permit cell mixing, rather that molecular influences are the critical determinants. Alkaline phosphatase staining of *Dh* mutant mice (see Introduction) reveals that the sharp expression boundary evident at the posterior stomach is maintained despite the absence of a pyloric sphincter (R. Watson; unpublished data).

5.3.2 Absence of *Shh* does not affect the *Bapx1* or *Nkx2.5* expression domains

Assessment of *Bapx1* and *Nkx2.5* expression in the developing gut of mice lacking *Shh* reveals that neither are affected by the loss of *Shh*. Previous studies have

suggested that SHH induces *Bmp4* expression in the small intestine which in turn acts through the *Bmp* receptor BMPR1B to upregulate *Nkx2.5* in a spatially restricted manner, the latter specifying the pyloric sphincter (Smith *et al.*, 2000a). The fact then that *Nkx2.5* is not lost in the absence of *Shh* suggests that, if these molecules reside in the same signalling pathway, another signalling factor (or factors) must be able to compensate for the loss of *Shh*. Perhaps this compensatory role could be accomplished by *Indian hedgehog* (*Ihh*) which has been found to be coexpressed with *Shh* only in the gut epithelium (Bitgood and McMahon 1995; Roberts *et al.*, 1995; Narita *et al.*, 1998). Additional studies have led to the proposition that *Shh* and *Ihh* may share the same or very similar signalling pathways (reviewed by Hammerschmidt *et al.*, 1997). Although these two genes reportedly have different roles in gut epithelial differentiation (Fukuda *et al.*, 2003) and different phenotypes are seen in the stomach and intestine of *Shh* and *Ihh* mutant mice (Ramalho-Santos *et al.*, 2000), this does not negate the possibility that some degree of functional compensation may occur between these hedgehog genes. The existence of redundancy amongst signalling molecules involved in patterning of the gut would not be surprising given the functional importance of this organ to survival. It further alludes to the wealth of signalling factors likely to be involved and the ensuing complexity of foregut patterning.

5.3.3 Mesenchymal gut markers in the *Bapx1* mutant gut

In the present investigation I was unable to detect transcripts from *Bmp2*, *Bmp7* or *Bmp11* at appreciable levels within the foregut. *Bmp2* has been implicated as important in gland formation within the chick proventriculus yet its expression is detected for only a limited time window (Narita *et al.*, 2000). The short temporal expression pattern may thus account for *Bmp2* not being detected in the current study. Moreover, the expression level of *Bmp11* in the SMP is low (Hecksher-Sørensen, 2001). Taken together, these observations suggest that neither *Bmp2*, *Bmp7* nor *Bmp11* assume primary responsibility for the gastroduodenal malformation manifest in the *Bapx1* mutant gut. The fact that no alteration to the wildtype molecular boundaries delimited by *Barx1* and *Six2* is observed in the *Bapx1* mutant

gut would also seem to exclude these molecules as effectors of the morphological transformation. Thus the loss of the posterior stomach boundary does not necessarily disrupt all molecular boundaries in this region of the gut since the expression levels and domains of some transcription factors are not altered.

5.3.3.1 What of the relationship between *Bmp4* and *Bapx1*?

The gut-specific expression pattern of *Bmp4* does not differ in the absence of *Bapx1*, in contrast to its alteration in the skeletal region of *Bapx1* null mice (Tribioli and Lufkin, 1999). *Bmp4* is detected in the anterior stomach and duodenum of both wildtype and *Bapx1* mutant mice but notably remains absent from the posterior stomach of the mutant. The constancy of *Bmp4* expression domains between wildtype and mutant guts suggests that this molecule is not responsible for the gastroduodenal malformation in *Bapx1*-/- mice. Furthermore, it suggests that *Bmp4* is not dependent upon direct regulation by *Bapx1* in the gut since, if it was, one would expect to detect *Bmp4* transcripts in the posterior stomach of *Bapx1* null mice. The upregulation of *Bmp4* in the skeletal region of *Bapx1* mutants was partially attributed to the observed downregulation of *Ihh* (Tribioli and Lufkin, 1999), thus losing the proposed inhibitory influence of *hedgehog* signalling upon *Bmp4* expression (Naski *et al.*, 1998). *Ihh* is normally localised to the posterior stomach of mice (Bitgood and McMahon, 1995), therefore if this expression domain was maintained in the *Bapx1* mutant it could account for the maintenance of *Bmp4* repression. Alternatively, active *hedgehog* signalling via other molecules could equally maintain *Bmp4* exclusion from the posterior stomach.

Although ectopic expression of *Bmp4* was reported in the chick gizzard (Nielsen *et al.*, 2001), it was not *consistently* observed and could possibly be attributed to the variant manipulation of *Bapx1* in the chick versus the mouse mutant models. Gene knockouts, widely used in the mouse including in the production of *Bapx1* null mutants, are not presently possible in the chick due to technical difficulties in the manipulation of one cell-stage embryos. Hence investigation of the role of *Bapx1* in the chick involved viral misexpression of a "reverse-function"

Bapx1 (Nielsen *et al.*, 2001). This may be a significant factor accounting for the difference between *Bmp4* expression observed in the *Bapx1* mutant mouse as compared to the "reverse-function" *Bapx1* construct in the chick. It is equally possible that evolutionary differences between the chick and mouse may account for the discrepancy discussed above. Indeed the gizzard is an avian-specific adaptation and has no direct mammalian counterpart, which may be an important consideration. It is also possible that regulatory differences exist between mammals and birds. The accumulation of *Bapx1* transcripts on opposite sides of the LPM in early mouse and chick embryos (Schneider *et al.*, 1999) is one allusion to the fact that, whilst the molecules themselves may be conserved, they may have evolved slightly different roles and be regulated in a marginally different manner in divergent species. Further examples exist of conserved factors that exhibit variant expression profiles in the mouse and chick guts, for example the transcription factors *Tcf4* and *Lef1*, which are downstream components of the β-catenin signalling pathway (Theodosiou and Tabin, 2003). Furthermore, recent investigations into the role of *Bapx1* in the murine middle ear support the proposal that, during evolution, the downstream targets of *Bapx1* in certain signalling cascades may have altered (see Introduction). Nevertheless, one of the conclusions from work in the chick was that other factors besides *Bapx1* are likely to play a role in *Bmp4* and *Wnt5a* regulation (Nielsen *et al.*, 2001), which is consistent with the data presented above.

5.3.4 Future directions

To further investigate the specification of molecular boundaries within the foregut and their disruption in the *Bapx1* null background will necessitate the identification of additional markers of specific domains along the A-P axis of the gut tube. In this respect, a comprehensive expression screen of *Wnt* signalling during different stages of gut development was recently undertaken in the chick (Theodosiou and Tabin, 2003). The initial overview of transcript distribution revealed that 12 of the 16 *Wnt* ligands analysed were expressed during various stages of gut development. Moreover, the expression profiles revealed these genes to be confined to distinct regions of the developing gut; transcript distribution was found to coincide neatly

with morphological boundaries. Thus, together with indications that expression domains of conserved molecules between mouse and chick are also generally conserved (for example consider the wildtype expression domains of *Bapx1*, *Shh*, *Bmp4* and *Wnt5a* in the gastrointestinal tract, further supported by evolutionary comparisons of molecular gut markers; Smith *et al.*, 2000b), carefully selected *Wnts* would be likely to prove beneficial in future investigations of boundary specification within the *Bapx1* mutant gut.

Besides the analysis of additional genes expressed in specific domains of the gut, it may prove useful to temporally extend the expression analysis of genes investigated in the present study. The definition of temporal limitations is likely to prove as important as the delineation of spatial domains to our understanding of normal gut development. The major challenge, however, will be to determine the way in which these patterning molecules interact with each other as it is their combinatorial effect that is ultimately responsible for the patterning of the gut. Establishment of the disrupted signal or signals responsible for the malformation of *Bapx1*-/- gut tissue described here would provide an important clue as to the molecular pathways involved in normal gut specification. Notably, the evolutionary conservation of these signalling factors and their associated pathways, together with their function in various anatomical domains, emphasises the importance of gut organogenesis as a model system for understanding development.

Chapter 6 Morphological development of the spleen

6 Morphological development of the spleen

6.1 Introduction

One of the phenotypic manifestations of loss of functional Bapx1 is asplenia, however the role of *Bapx1* in the formation of the spleen has received little attention. Perhaps this is a reflection of the dearth of information regarding the specification, patterning and early morphogenesis of this organ. Historically, the literature pertaining to the spleen focuses principally on its early role in haematopoiesis and later on immunology-associated functions. Thus detailed anatomical and histological descriptions exist that essentially relate to developmental stages at which this organ is already well-established morphologically and has adopted its ultimate position attached to the dorsal mesogastrium flanking the stomach. The precise origin of the spleen, however, and its initial morphogenesis remain poorly chronicled. This is in contrast to the relative abundance of literature regarding the development and patterning of the stomach and pancreas which develop in close association with the spleen. Indeed the dorsal pancreatic bud is flanked by spleno-pancreatic mesenchyme, the presence of which is essential for pancreas development (Golosow and Grobstein, 1962) and within which the spleen develops, yet despite this intimate association, detailed descriptions of pancreas development largely ignore the existence of the spleen.

Previous work in our lab led to the characterisation of a novel anatomical structure referred to as the SMP which is hypothesised to play a role in the leftward growth of the spleno-pancreatic mesenchyme and which is defective in the absence of *Bapx1* expression (Hecksher-Sørensen *et al.*, 2003, submitted; see Introduction). The splenic markers *Hox11* and *Nkx2.5*, expressed at E10.5 in the splanchnic mesenchyme underlying the SMP, were shown to be absent in *Bapx1*-/- embryos, demonstrating that these two genes lie downstream of *Bapx1* in the spleen induction pathway (Hecksher-Sørensen *et al.*, 2003, submitted). This work highlighted the location along the anterior-posterior (A-P) axis of the foregut at which spleen development begins, namely in the vicinity of the dorsal pancreatic bud posterior to the stomach, and thus paved the way for the current evaluation of early spleen morphogenesis. The present study was greatly facilitated by the recent development within the Unit of a

novel tool for gene expression studies, Optical Projection Tomography (OPT) (Sharpe *et al.*, 2002; Sharpe, 2003). Notably, this advanced imaging technique allows spleen development to be followed within the 3 dimensional (3D) context of the surrounding gut tissue.

6.2 Results

Embryos were harvested at a number of different developmental stages and, since a more accurate classification of the precise developmental stage was deemed necessary for the present study, they were subsequently staged and grouped by somite number (see Materials and Methods). The relationship between somite number and embryonic day (E) during early mouse development is shown in Table 6.1. The guts were dissected from each embryo and processed for further analysis. Spleen tissue was identified via expression of the spleen-specific marker *Hox11* as detected by wholemount *in situ* hybridisation. Whilst *Bapx1* and *Nkx2.5* also constitute splenic markers, only *Hox11* is limited to the spleen within the foregut region.

Somites	Mouse (Embryonic day)
8-14	8.5-9.0
13-20	9.0-9.5
21-29	9.5-10.25
30-39	10.25-10.75
40-45	11.0-11.5

Table 6.1: A table detailing the relationship between somite number and embryonic day (E) in the developing mouse embryo.

6.2.1 A time course of spleen development

The data presented in this chapter is an amalgamation of imaging both by conventional means and by OPT. The five representative developmental stages selected to illustrate spleen morphogenesis correspond to 36-38 somites (E10.5), 43-44 somites (E11.0), 48-49 somites (E11.5), 51-52 somites (E12.0) and 57-58 somites (E12.5). Guts dissected from 28 somite embryos (E10.0) were also analysed however a specific *Hox11* signal corresponding to splenic precursor cells was not detected at this early stage. Nevertheless, a 28 somite gut is included in the representative time course of spleen development collectively presented in Figure 6.1. A dense mass of cells is visible posterior to the stomach along the A-P axis and this domain, by comparison to the E10.5 foregut, is judged to be the location of the developing dorsal pancreas, and thus by extrapolation, the position from which the spleen will develop (Figure 6.1A). The marked morphogenesis of the spleen, involving extensive growth and elongation, anterior movement and asymmetric positioning is evident within a relatively limited developmental window (E10.5-E12.5) (Figure 6.1B-F). 3D rotating movies were produced of the foregut at 28, 36, 43, 49 and 58 somite stages and are included in the attached CD. Each of the five stages analysed in detail is described below.

6.2.1.1 36-38 somites

The schematic in Figure 6.2A enables familiarisation with the principal features and arrangement of the foregut region at E10.5. At this stage, *Hox11* is faintly detected in whole guts posterior to the stomach, localised to tissue positioned lateral to the anterior portion of the duodenum (Figure 6.2B, C, E, F). This tissue includes the dorsal pancreatic bud and surrounding splanchnic mesenchyme. *Hox11* transcripts are more readily apparent in transverse sections taken through the foregut at the level of the dorsal pancreatic bud (Figure 6.2K and L). *Hox11* expression is seen underlying the thickened mesothelial layer of the SMP in the mesenchyme on the dorsal side of the dorsal pancreatic endoderm. Thus by the 36 somite stage, at which point splenic precursor cells are detectable, they are already asymmetrically positioned relative to the embryonic midline and analysis of whole guts suggests that

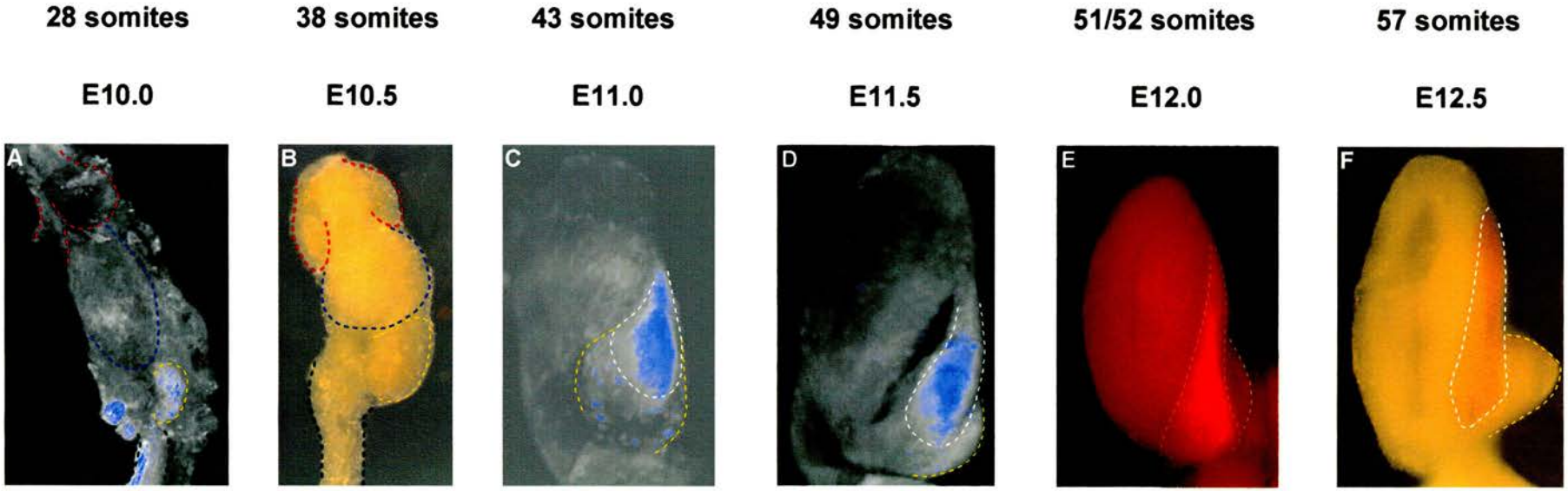


Figure 6.1: An overview of spleen development as illustrated by six images selected to represent key stages during splenogenesis. At E10.0 (**A**), *Hox11*-expressing splenic precursor cells are not detected however the dense patch of tissue outlined in yellow, located posterior to the stomach (outlined in blue), marks the approximate location of the dorsal pancreatic bud which is surrounded by mesenchyme from which the spleen will develop. The lungbuds and duodenum are outlined by red and green dotted lines respectively. *Hox11* is detectable by E10.5 (**B**) in the dorsal pancreatic mesenchyme surrounding the pancreatic bud which is growing in an anterior direction towards the posterior stomach. The growth and anterior movement of the spleno-pancreatic structure towards an asymmetric position on the right side of the stomach is evident by E11.0 (**C**). At this stage, *Hox11* signal (which marks splenic tissue and is outlined in white and pseudo-coloured blue) is localised in the anterior portion of the spleno-pancreatic structure. Anterior extension of the spleen proceeds (**D**), possibly aided by 'anchors' at the anterior and posterior tips of the spleno-pancreatic structure (**C**, **D**). The morphological distinction between the spleen and the pancreas is evident by E11.5 (**D**). By E12.0, the spleen has developed into a long, narrow structure and has essentially reached its final position attached to the dorsal mesogastrium flanking the anterior stomach (**E**, **F**). The dorsal pancreas adopts a more pointed appearance and is located ventral to the spleen, around its posterior end along the A-P axis (**F**). The *Hox11* signal is detected by a Fast Red *in situ* probe. **A**, **C**, **D**, foreguts analysed by OPT; **B**, **F**, brightfield photography; **E**, texas red fluorescence imaging. Refractive light effects resulting from the necessity of scanning OPT specimens in glycerol account for non-specific signal visible in images **A**, **C** and **D**.

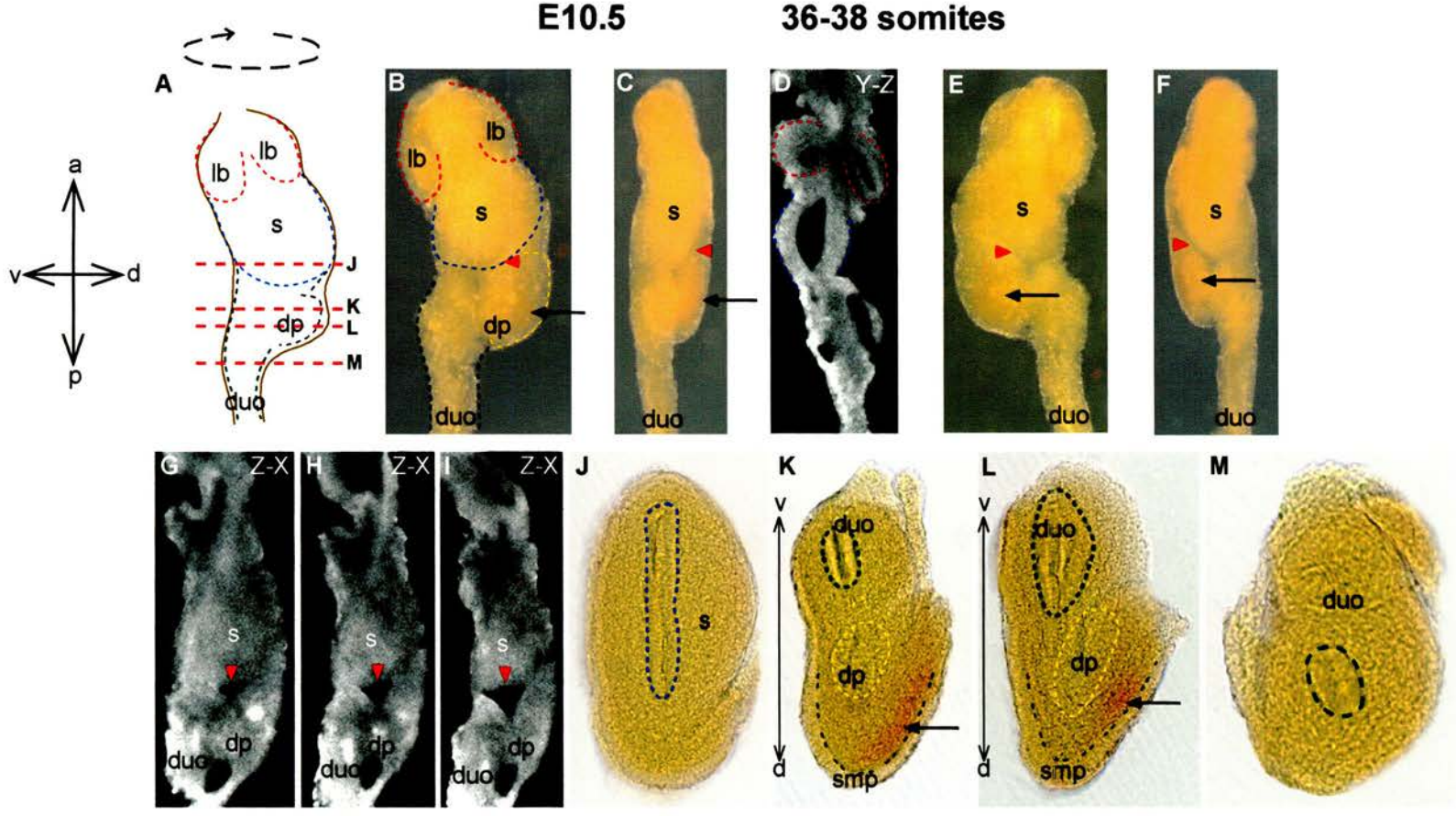


Figure 6.2: Localisation of the developing spleen at E10.5 (36-38 somites) by wholemount *in situ* hybridisation using a *Hox11* Fast Red *in situ* probe. **(A)** Schematic to show the relative positions of organs within the foregut. Rotation of the gut clockwise on its perpendicular axis is indicated and relates to the successive positions at which guts appear in **B** (left lateral view) through **E** (right lateral view) to **F**. *Hox11* is detected in the dorsal pancreatic mesenchyme (black arrow), marking the position of the developing spleen posterior to the developing stomach, as demonstrated by transverse sections (**J-M**) through the foregut at the A-P level indicated in **A**. *Hox11* is detected in cells underlying the splanchnic mesothelial plate (smp) at the level of the dorsal pancreatic bud (**K, L**), but is not detected in the stomach (**J**) or the duodenum (**M**). A distinction can be made between the structure of the stomach and the developing spleno-pancreatic region below, due to thinning of the mesenchyme in the intervening region (red arrowhead). A difference in cell density along the A-P axis is also evident in consecutive virtual sagittal sections through a gut analysed by OPT (**G-I**). **(D)** Virtual frontal section through the foregut, as seen by OPT, to aid localisation of the lungbuds (outlined in red) and stomach (outlined in blue) in flanking images (**B**, **C**, **E**, **F**). a, anterior; d, dorsal; dp, dorsal pancreas; duo, duodenum; lb, lungbud; p, posterior; s, stomach; smp, splanchnic mesothelial plate; v, ventral. The orthogonal plane of OPT sections **D** and **G-I** is indicated to the top right of each image.

anterior growth towards the posterior stomach is underway (Figure 6.2C, F). Views of the foregut tissue from various angles demonstrate that the *Hox11*-expressing tissue is clearly independent of the posterior stomach, being separated by a region of less dense mesenchymal tissue (Figure 6.2C, F). The difference in cell density defining the posterior stomach as distinct from the spleno-pancreatic tissue is also evident in virtual sagittal sections through a 36 somite gut analysed by OPT (Figure 6.2G-I). In the absence of molecular markers, the spleen and dorsal pancreas are morphologically indistinguishable at the 36-38 somite stage.

6.2.1.2 43-44 somites

By this stage, whilst the spleen and dorsal pancreas still constitute a continuous aggregation of tissue, this tissue has acquired a more definitive structure (Figure 6.3B). OPT analysis demonstrates that the domain of *Hox11*-expressing cells is clearly more extensive than at the 36-38 somite stage and is confined to the anterior portion of the spleno-pancreatic tissue (Figure 6.3C). The dorsal pancreas does not express *Hox11* and occupies a more posterior position to that of the spleen. The distinct anterior domain of *Hox11* expression is also evident in virtual frontal sections of the foregut obtained by OPT analysis (Figure 6.3F-I). As opposed to its initial position posterior to the stomach, the spleen has moved in an anterior direction such that it flanks the caudal end of the developing stomach. In terms of positioning along the A-P axis, the anterior *Hox11*-expressing domain appears to be approximately halfway along the length of the stomach wall (Figure 6.3B, C, E), alluding to a rapid rate of growth and elongation between E10.5 and E11.0. The spleno-pancreatic tissue appears to be growing asymmetrically along the right lateral side of the dorsal stomach wall as opposed to remaining along the midline (Figure 6.3B-I). Asymmetric positioning on the righthand side is also evident in virtual transverse OPT sections (Figure 6.3J).

The thinning of tissue marking the separation between the developing stomach and spleno-pancreatic organs, evident at the 36-38 somite stage, is again apparent (Figure 6.3B-D, F-I). By the 43-44 somite stage, however, the anterior and posterior tips of the spleno-pancreatic structure appear to be anchored to the adjacent stomach wall, forming a continuum, whilst it is the intervening region that appears to be

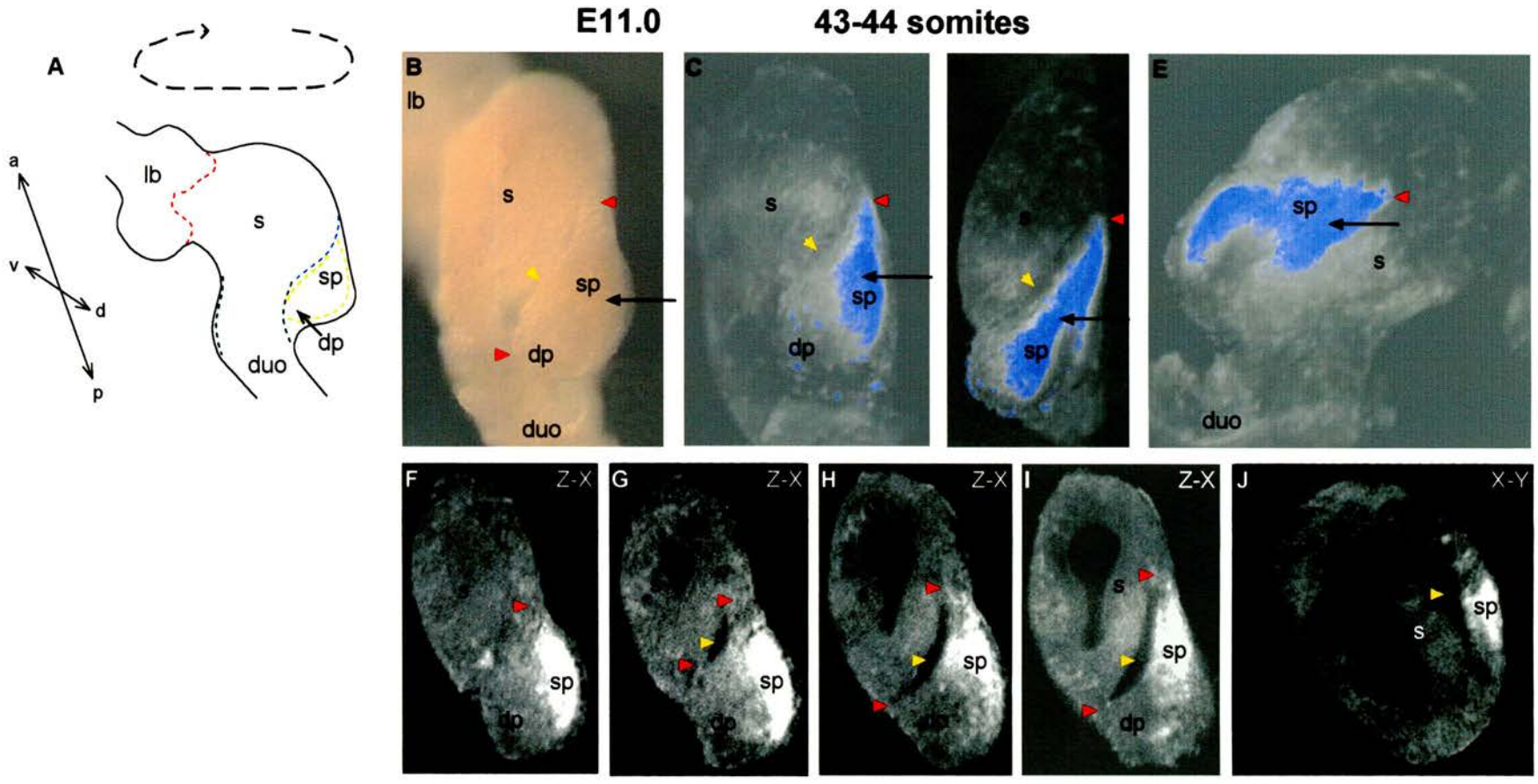


Figure 6.3: Localisation of the developing spleen at E11.0 (43-44 somites) by wholemount *in situ* hybridisation using a *Hox11* probe. (**A**) Schematic to illustrate the position of the developing spleen relative to the posterior stomach and dorsal pancreas as seen from a left lateral perspective. Rotation 90° clockwise of the gut in **A** gives a dorsal view as shown in **B** and by OPT in **C**. The *Hox11* signal is localised to the anterior half of the structure encompassing both the developing spleen and dorsal pancreas, as seen in wholemount guts (**B**, **C**) and in consecutive virtual frontal sections in a dorsal to ventral direction generated by OPT analysis (**F-I**). Between 38 and 43 somites, the spleno-pancreatic structure has increased in size and moved in an anterior direction to lie adjacent to the posterior stomach (**B**, **C**). The spleen extends upwards adjacent to the right lateral wall of the stomach rather than following the dorsal midline (**E**). There appear to be two principal 'anchors' connecting the spleno-pancreatic structure to the adjacent tissue (red arrowheads in **B-I**), with intervening tissue being less dense (yellow arrowheads in **B-D**, **G-J**). The *Hox11* expression domain narrows towards the tip at the anterior 'anchor' as shown by OPT analysis (**C-E**) whereby the gut in **C** was rotated clockwise around the dorsal-ventral axis. Virtual transverse sections detect *Hox11* signal in the spleen but not in the adjacent stomach tissue (**J**). a, anterior; d, dorsal; dp, dorsal pancreas; duo, duodenum; lb, lungbuds; p, posterior; s, stomach; sp, spleen; v, ventral. The orthogonal plane of OPT sections **F-J** is indicated to the top right of each image. The fluorescence signal detected by OPT in wholemount guts (**C-E**) is pseudo-coloured blue.

composed of less dense tissue. This observation is evident in whole guts imaged both by conventional means (Figure 6.3B) and by OPT (Figure 6.3C, D). Furthermore, the anterior and posterior 'anchors' are also evident in frontal sections (Figure 6.3F-I) and transverse sections (Figure 6.3J) of foreguts analysed by OPT. The latter imaging technique enabled the tissue to be viewed from any angle and thus facilitated a more detailed analysis. Rotation of the stomach around the dorsal-ventral (D-V) axis permitted a closer inspection of the anterior domain of the developing spleen. The *Hox11* expression domain progressively narrows towards its most anterior tip (Figure 6.3D, E), creating the appearance of a 'leading edge' in the anterior extension of the spleen domain. This 'leading edge' appears to be approximately coincident with the anterior 'anchor' which, in turn, seems to form a tight connection to the stomach wall and is potentially required to support the elongation of the spleen.

6.2.1.3 48-49 somites

Elongation has evidently persisted between the developmental stages of E11.0 and E11.5, resulting in further extension of the spleen in an anterior direction along the dorsal, right lateral stomach wall. The anterior 'leading edge' of the spleen extends beyond halfway along the A-P axis of the stomach in a posterior-anterior direction (Figure 6.4A-D, F) and the posterior domain of the spleen now resides anterior to the pyloric sphincter (Figure 6.4A, F, I). Analysis of the foregut from a variety of different angles demonstrates that the spleen and dorsal pancreas are beginning to be distinguished morphologically as distinct organs by this stage. The *Hox11*-expressing spleen tissue is starting to resemble its mature form in that it appears as a narrow, elongated structure, in contrast to the dorsal pancreas, which has a more rounded appearance, occupies a smaller area and retains a position towards the posterior end of the developing spleen (Figure 6.4C-F). The *Hox11* expression domain at 48-49 somites is more extensive than at 43-44 somites, as seen from a dorsal perspective (Figure 6.4C) and in virtual frontal sections (Figure 6.4E). Whilst the principal growth of the spleen occurs along the A-P axis, the position of the dorsal pancreas has not significantly altered along the A-P axis, however lateral growth is apparent, resulting in its emerging to the right, and ventral to the spleen (Figure 6.4C-F).

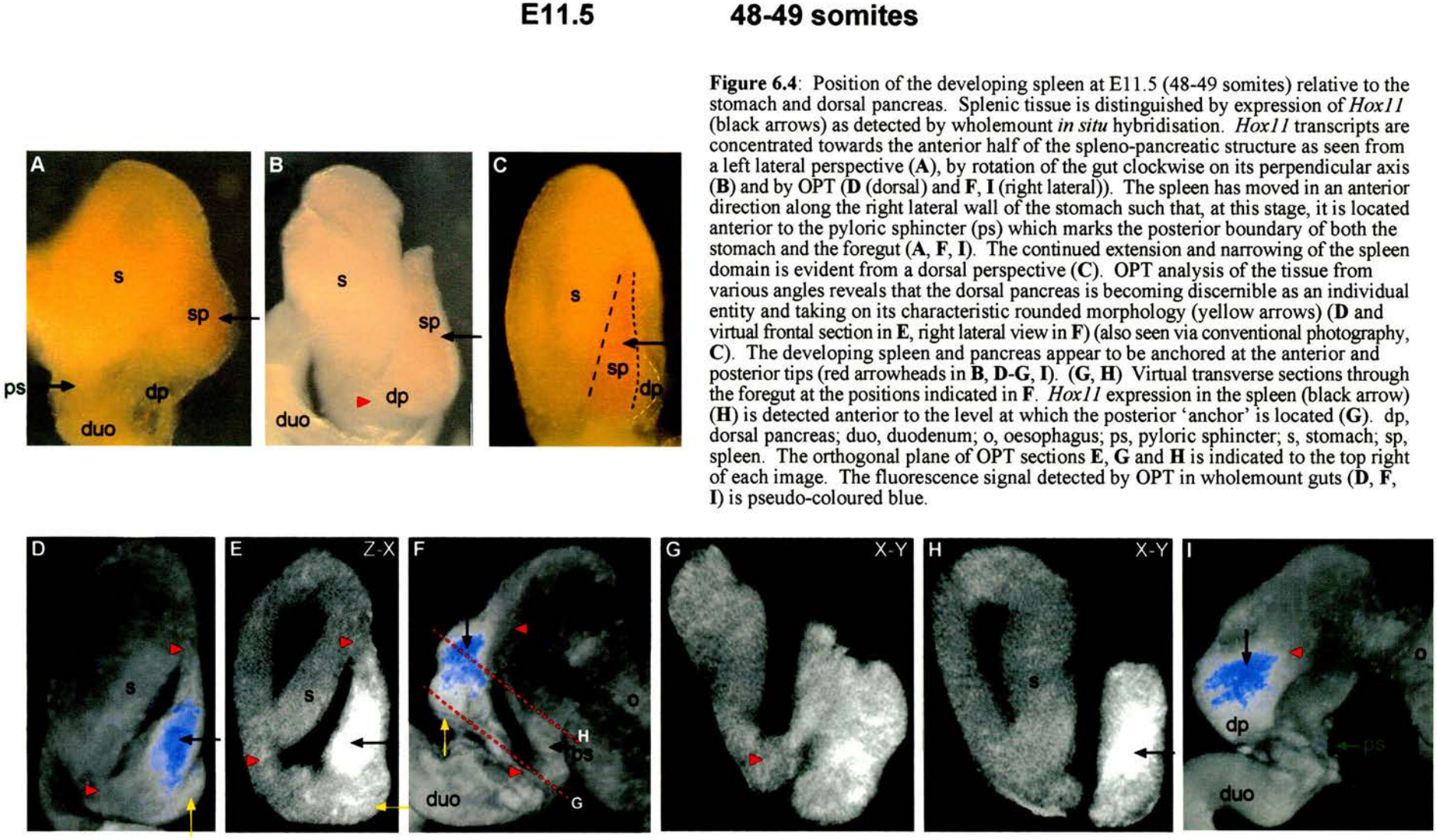


Figure 6.4: Position of the developing spleen at E11.5 (48-49 somites) relative to the stomach and dorsal pancreas. Splenic tissue is distinguished by expression of *Hox11* (black arrows) as detected by wholemount *in situ* hybridisation. *Hox11* transcripts are concentrated towards the anterior half of the spleno-pancreatic structure as seen from a left lateral perspective (**A**), by rotation of the gut clockwise on its perpendicular axis (**B**) and by OPT (**D** (dorsal) and **F**, **I** (right lateral)). The spleen has moved in an anterior direction along the right lateral wall of the stomach such that, at this stage, it is located anterior to the pyloric sphincter (ps) which marks the posterior boundary of both the stomach and the foregut (**A**, **F**, **I**). The continued extension and narrowing of the spleen domain is evident from a dorsal perspective (**C**). OPT analysis of the tissue from various angles reveals that the dorsal pancreas is becoming discernible as an individual entity and taking on its characteristic rounded morphology (yellow arrows) (**D** and virtual frontal section in **E**, right lateral view in **F**) (also seen via conventional photography, **C**). The developing spleen and pancreas appear to be anchored at the anterior and posterior tips (red arrowheads in **B**, **D-G**, **I**). (**G**, **H**) Virtual transverse sections through the foregut at the positions indicated in **F**. *Hox11* expression in the spleen (black arrow) (**H**) is detected anterior to the level at which the posterior 'anchor' is located (**G**). dp, dorsal pancreas; duo, duodenum; o, oesophagus; ps, pyloric sphincter; s, stomach; sp, spleen. The orthogonal plane of OPT sections **E**, **G** and **H** is indicated to the top right of each image. The fluorescence signal detected by OPT in wholemount guts (**D**, **F**, **I**) is pseudo-coloured blue.

As at the 43-44 somite stage, the spleno-pancreatic tissue appears to possess anterior and posterior 'anchors' to the adjacent tissue. The intervening tissue, meanwhile, is considerably less dense and OPT analysis from various angles suggests that it is physically separate (Figure 6.4D, F). Potential physical separation of intervening tissue between the anterior and posterior 'anchors' is also suggested by both virtual frontal (Figure 6.4E) and transverse (Figure 6.4H) sections. The posterior connection of spleno-pancreatic tissue was identified by the inspection of serial virtual transverse sections through the foregut (Figure 6.4G) whereby sections anterior to the proposed 'anchor', at the level of the developing spleen, detect the stomach and *Hox11*-expressing spleen as separate entities (Figure 6.4H). At the 48-49 somite stage, as the organs become more distinct, the location of the posterior 'anchor' is more precisely determined. Whilst the anterior connection, as at E11.0, seems to be between the 'leading edge' of the spleen and the adjacent dorsal wall of the stomach (Figure 6.4A, B, D-F, I), the posterior connection appears to be to duodenal tissue posterior to the level of the pyloric sphincter (Figure 6.4F). This posterior 'anchor', as at E11.0, does not involve *Hox11*-expressing spleen tissue, which has migrated in an anterior direction along the A-P axis.

6.2.1.4 51-52 somites

Elongation of the spleen in an anterior direction flanking the right dorsal wall of the developing stomach persists through the 51-52 somite stage as demonstrated by the extensive *Hox11*-expressing domain (Figure 6.5A). Transverse sections demonstrate that only flanking the upper-most region of the anterior stomach is *Hox11* not expressed (Figure 6.5A), alluding to the continued anterior migration of splenic tissue. Transverse sections taken approximately half way along the A-P axis of the stomach illustrate the strong and specific expression of *Hox11* in the spleen, which at this A-P level lies immediately lateral to the stomach wall (Figure 6.5C). At a more posterior level, the spleen lies dorsal, as opposed to lateral, with respect to the stomach and *Hox11* expression is detected around the outer region of tissue that appears as a bulge at the level of the dorsal pancreas (Figure 6.5D, E). The expression pattern reflects the observation that the spleen now lies as an elongated, narrow structure that curves

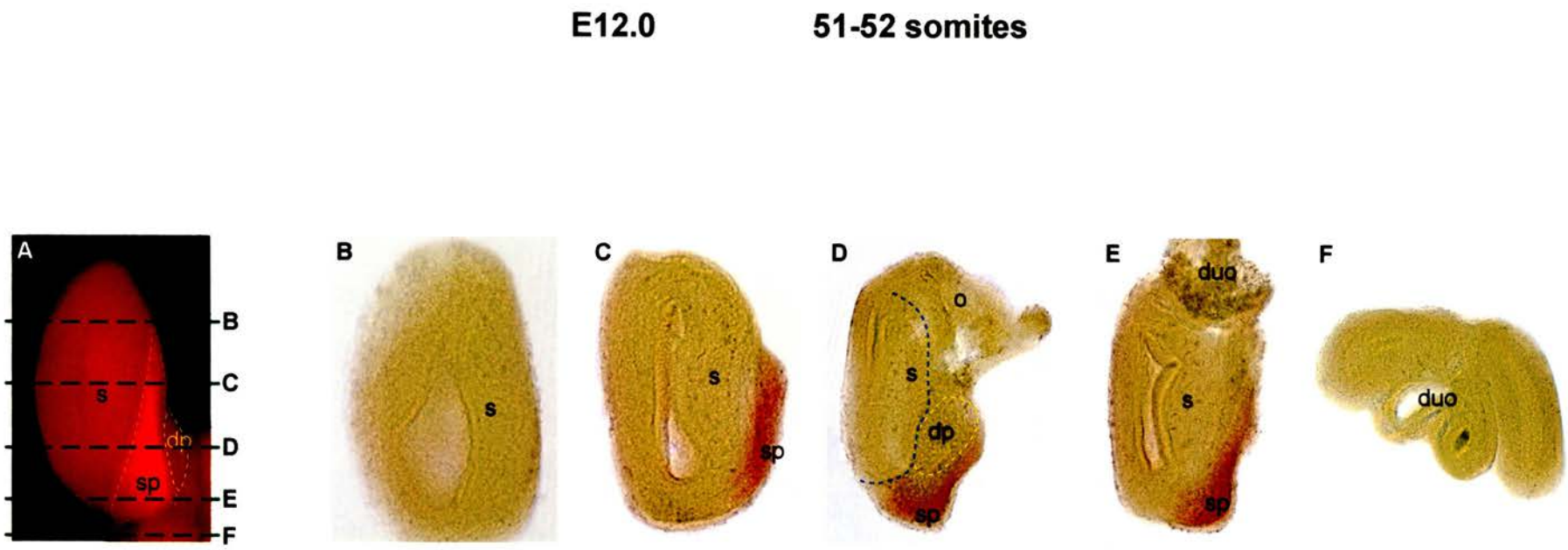


Figure 6.5: Position and morphology of the spleen at E12.0 (51-52 somites). *Hox11*-expressing spleen tissue is detected by wholemount *in situ* hybridisation of dissected guts using a *Hox11* RNA *in situ* probe. (**A**) The characteristic triangular shape of the spleen at mid-gestation as visualised by texas red fluorescence from a dorsal perspective. The spleen domain has extended in an anterior direction along the dorsal wall of the stomach whilst the dorsal pancreas, which does not express *Hox11*, has a more rounded morphology than the spleen, does not extend to flank the anterior stomach and grows away from the stomach wall rather than alongside it. Green dotted lines in **A** relate to the corresponding transverse sections in **B-F**. *Hox11* expression is detected both anterior (**C**) and posterior (**E**), in addition to at the level of (but not within), the developing dorsal pancreas (**D**). dp, dorsal pancreas; duo, duodenum; o, oesophagus; s, stomach; sp, spleen.

around the upper wall of the stomach. Guts at the 51-52 somite stage were not analysed by OPT.

6.2.1.5 58 somites

Little significant alteration to the morphology of the spleen is detected between the 51-52 somite and 58 somite stages of embryonic development by analysis of *Hox11* expression. By E12.5, the spleen has essentially adopted its ultimate position with respect to the stomach and dorsal pancreas, as demonstrated by comparison to the E14.5 foregut (Figure 6.6A, I). The dorsal pancreas lies ventral to the spleen (Figure 6.6E-G) and, by E12.5, appears as a slightly more cuspidate structure (Figure 6.6A, B, G). Transverse sections reveal that the dorsal pancreas abuts the stomach along a considerable portion of the D-V axis (Figure 6.6H), in contrast to the spleen, which covers a greater length along the A-P axis. This observation was clearly demonstrated by removal of the dorsal mesogastrium, to which the spleen and dorsal pancreas are attached, at E14.5 (Figure 6.6I). There is no overlap at this stage between the attachment of the spleen and dorsal pancreas to the dorsal mesogastrium. As at E12.5, the dorsal pancreas resides in the vicinity of the posterior end of the spleen (Figure 6.6A, I).

In contrast to earlier developmental stages, for example the 43-44 somite stage, at which time anterior migration of *Hox11*-expressing tissue is underway and there appears to be a 'leading' anterior edge to the spleen, at the 58 somite stage, the anterior domain appears to have levelled such that the spleen takes on an almost rectangular morphology (Figure 6.6C). Furthermore, the anterior 'anchor', detectable until the 48-49 somite stage, is not visible in the 58 somite gut, at which point elongation is essentially complete. Rather, the attachment appears to be uniform along the length of the spleen and the flanking wall of the stomach (Figure 6.6A-C, E, F). The posterior 'anchor', however, is still detectable at the 58 somite stage (Figure 6.6D). It appears to be between duodenal tissue posterior to the pyloric sphincter and non-*Hox11* expressing tissue that initiates at the A-P level in the vicinity of the dorsal pancreas. The intervening tissue appears to be either considerably less dense, or is perhaps, at least in part, physically separated (Figure 6.6D, F-H). At E14.5, removal of the dorsal mesogastrium, which is a sheath-like covering of the stomach wall,

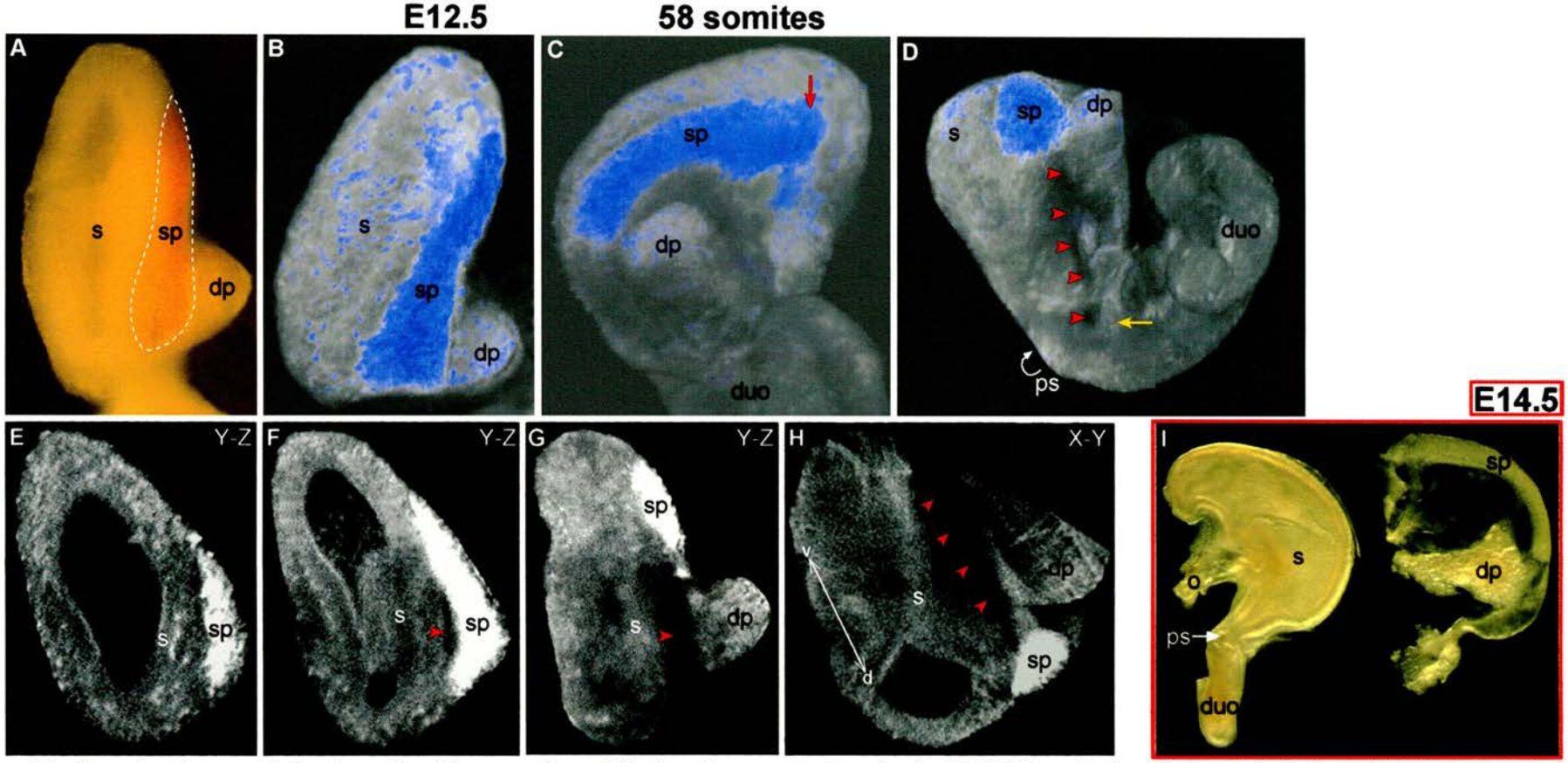


Figure 6.6: The spleen has essentially adopted its ultimate position within the spleno-pancreatic region by E12.5 (58 somites) and is morphologically well-defined. Splenic tissue is distinguished by expression of *Hox11* as detected by wholemount *in situ* hybridisation of dissected guts (**A-H**). (**A**, **B**) Dorsal and (**C**) right lateral view. The spleen extends along the dorsal, right lateral wall of the stomach and is a narrow elongated structure as compared to the shorter, more pointed pancreas (**A**, **B**, **G**). Both virtual frontal sections (**E-G**), from a more dorsal (**E**) to a more ventral (**G**) position, and virtual transverse sections (**H**) demonstrate the position of the spleen relative to the dorsal pancreas at E12.5. The latter occupies a more ventral location than the spleen and abuts the stomach along the D-V axis (**H**) whilst remaining at approximately the same position along the A-P axis (also evident in **I**). OPT analysis of guts from a right lateral perspective (**C**) demonstrates that the anterior domain of the spleen (red arrow) has levelled such that the structure has taken on an almost rectangular morphology, curved asymmetrically around the stomach wall. The anterior 'anchor' is not detected at E12.5; at this stage the attachment between the spleen and the stomach appears to be uniform along the length of the spleen (**A-C**, **E**, **F**, **I**). There does, however, appear to be a posterior connection maintained between a non-*Hox11*-expressing region emanating posterior to the dorsal pancreas and tissue posterior to the pyloric sphincter (yellow arrow) (**D**). In this posterior half of the foregut, beginning at the level of the dorsal pancreas and ending at the posterior 'anchor', the intervening tissue is either less dense or physically separated (red arrowheads) (**D**, **F-H**). (**I**) E14.5 dissected gut (left) with the dorsal mesogastrium stripped away from the wall of the stomach (right). The spleen and dorsal pancreas are attached to the dorsal mesogastrium with the spleen flanking primarily the anterior stomach. The posterior part of the spleen lies in the middle of the A-P region occupied by the dorsal pancreas, as seen also at E12.5 (**A**). There is no marked alteration to either spleen morphology or position between E12.5 and E14.5. d, dorsal; dp, dorsal pancreas; duo, duodenum; o, oesophagus; ps, pyloric sphincter; s, stomach; sp, spleen; v, ventral. The orthogonal plane of the OPT section in **E-H** is indicated to the top right of each image. The fluorescence signal detected by OPT in wholemount guts (**B-D**) is pseudo-coloured blue. Refractive light effects resulting from the necessity of scanning OPT specimens in glycerol account for non-specific signal visible in images **B-D**.

illustrates that, by this stage, the spleen and dorsal pancreas are physically separate organs (Figure 6.6I).

6.3 Conclusions and Discussion

In the present chapter, expression of the spleen marker *Hox11* was detected by *in situ* hybridisation in dissected guts at five successive time points during embryogenesis between and inclusive of 36 and 58 somites. The astonishing rapidity of early spleen morphogenesis is evident in the remarkable transition from a small domain of *Hox11*-expressing cells posterior to the stomach to an extensive layer flanking the predominantly anterior, and right lateral, dorsal mesogastrium. One of the significant advantages of OPT imaging is that specimens can be viewed and images subsequently captured from any angle, together with the instant accessibility to virtual sections taken in the transverse, sagittal or frontal planes. This facilitates examination of specimens in considerable detail and in the present analysis, permitted visualisation of spleen development in a 3D context whereby cells contributing to the spleen were identified by their expression of the spleen-specific marker *Hox11* and were thus located relative to the associated foregut tissue.

6.3.1 Defining features of early spleen development

The principal morphological changes appear to take place between 36 and 49 somites. Within this developmental window there is a transformation from a dense mass of tissue having a nondescript shape and positioned posterior to the stomach, to a defined, elongated structure which has migrated in an anterior direction relative to the stomach and is asymmetrically juxtaposed to the stomach wall.

6.3.1.1 The existence of a splenic 'scaffold'?

Interestingly, by E11.0, processes referred to as 'anchors' were observed emanating from the anterior and posterior termini of the spleno-pancreatic structure which

appeared to form connections to the adjacent tissue. Images to illustrate these 'anchors' were captured both by conventional means, and following the processing of gut tissue for OPT analysis, suggesting that these observations were not merely artefacts of the OPT imaging technique. Both the anterior and posterior 'anchors' were detectable until E12.0, thus coinciding with the period of rapid elongation and anterior extension of the spleen. After E12.0, only the posterior 'anchor' was apparent. It thus appears that the anterior domain of the developing spleen and posterior (presumably pancreatic) tissue may be fused, and thus anchored, to the respective flanking tissue, perhaps to provide some form of supportive scaffold during the development and elongation process. The existence of a continuum between tissue at the anterior and posterior termini is further supported by the identification of intervening regions of non-continuous tissue which, rather than representing areas of lower cell density, may in fact reflect the physical separation of tissues. The size of the apparent gap appears to increase slightly in an anterior to posterior direction along the length of the elongating spleno-pancreatic structure. Furthermore, by E12.5, at which point only the posterior 'anchor' is detected, the tissue posterior to the level of the dorsal pancreas is more loosely attached to the flanking tissue than the spleen, which is uniform along its length in its attachment to the adjacent stomach wall. It is possible that, once the anterior elongation is essentially complete and the spleen has reached its ultimate position relative to the flanking foregut tissue, the anterior splenic 'anchor' is no longer required for support. During subsequent embryonic and postnatal development, the growth of the spleen to realise its adult length may reflect a proportional increase in the size of the stomach, whilst its relative position remains unaltered. This could explain the early requirement for a tight connection at the 'leading' terminus during the rapid elongation process, but its subsequently being surplus to developmental requirements.

6.3.1.2 The nature of the 'anchor' tissue

The aforementioned observations pose a number of intriguing questions, one of which pertains to the cellular identity of the tissue forming the 'anchor'. This structure could potentially be of endodermal origin, as in the case of the neighbouring stomach and duodenum, or of mesodermal origin, since the 'anchor' connects to the spleno-

pancreatic mesenchyme. It could equally constitute a region of mixed cell type. These questions could potentially be addressed by the detailed examination of somite stage-matched guts, employing various endodermal markers to distinguish the cellular identity of different tissues. In this respect, during finalisation of the experimental strategy, dual localisation of foregut markers was attempted for both PDX1/*Hox11* and E-Cadherin/*Hox11*. PDX1 marks the pancreatic endoderm whilst E-Cadherin allows visualisation of the entire foregut endoderm. Whilst both the antibodies and *in situ* probe worked well independently, numerous variations on a combined immunolocalisation/*in situ* hybridisation protocol indicated that they were not compatible. Nevertheless, preliminary results in the lab suggest that alkaline phosphatase, whilst becoming restricted to its classical intestinal pattern between E11.5-E13.5, is expressed throughout the foregut endoderm at E10.5 (R. Watson, unpublished data). The detection of this enzyme involves a simple staining reaction as opposed to immunolocalisation, thus alkaline phosphatase staining may function in conjunction with *Hox11 in situ* hybridisation. Such an investigation could be implemented to explore the early stages of spleen development with regard to both the nature of the 'anchor' tissue and to the provision of an orientation marker. Assessment of the contribution of stomach-specific tissue to the anterior splenic 'anchor' would likely be facilitated by the use of an endodermal probe expressed in the anterior stomach. To this end, double label *in situ* hybridisation using *Shh* combined with *Hox11* may be informative.

6.3.1.3 The acquisition of positional identity

Another fascinating issue relates to the patterning events responsible for the expression of *Hox11* in an anterior domain of the spleno-pancreatic tissue, and for the apparent existence of splenic 'anchors'. It is possible that, prior to the appearance of the anterior and posterior processes, the spleno-pancreatic tissue may already be patterned along the A-P axis such that cells within the dense splanchnic mesenchymal domain, although constituting a region of nondescript shape, have acquired a positional identity. In this way, proliferation could result in extensive anterior extension of the *Hox11*-expressing spleen domain, necessitating an anterior 'anchor' to direct growth and provide support, whilst the posterior region is limited in its

movement along the A-P axis, its ultimate location relative to the posterior foregut being pre-determined.

Interestingly, nested expression patterns within the gut endoderm have been reported for a variety of transcription factors such as *Hepatocyte Nuclear Factor 3* (*HNF3*) α, β and γ (Monaghan *et al.*, 1993) and numerous members of the *Wnt* family of signalling molecules (Theodosiou and Tabin, 2003). Furthermore, analysis of the SMP has demonstrated differential expression of the transcription factors *Fgf9* and *Fgf10* in the dorsal and ventral domains of the SMP respectively, together with expression of *FgfR3*, a receptor for *Fgf9*, also limited to the dorsal SMP (Hecksher-Sørensen *et al.*, 2003, submitted). Given the putative involvement of such genes in the establishment of an epigenetic code for patterning different regions of the gut tube, it is plausible that the acquisition of spatial awareness is an integral part of early spleen morphogenesis. The molecules responsible remain to be elucidated, and they are likely to reside upstream of *Bapx1* in the spleen development pathway since the latter is expressed throughout the spleen, as are its downstream targets *Hox11* and *Nkx2.5*.

6.3.2 Two models to explain spleen development

Between 20-25 somites, mesenchyme accumulates around the epithelium of the dorsal pancreas (Wessells and Cohen, 1967; Pictet *et al.*, 1972), separating the presumptive pancreatic endoderm from the dorsal aorta. By this stage a dorsal and a ventral pancreatic bud have formed from the dorsal gut endoderm. Cells within the condensed region of spleno-pancreatic mesenchyme destined to become spleen are similarly envisaged to have been specified by an as yet unidentified patterning signal. Prior to E10.5, however, *Hox11* expression in splenic precursor cells is not detectable. Rapid proliferation to increase cell density takes place with minimal concomitant movement of splanchnic mesenchyme until the ~E10.5 (36 somite) stage. At this point, the SMP could initiate significant asymmetric mesenchymal growth, perhaps by the local provision of growth factors, such that the splanchnic mesenchyme extends rapidly in an anterior direction towards the caudal end of the stomach. The direction

and nature of the movement may also be influenced by physical space constraints within the developing foregut.

SMP-derived and/or other signals must also account for differentiation such that the tissue has acquired a more defined structure by the 43 somite stage. On the other hand, tissue differentiation could be accomplished by *Hox11* expression, which is limited to cells contributing to the spleen and for which the initial detection of expression directly precedes the principal morphological changes. Expression is subsequently maintained throughout the stages of spleen morphogenesis described here and it has previously been suggested that *Hox11* is essential for cell survival during spleen development (Dear *et al.*, 1995).

Amalgamating the results described thus far, together with previous data, leads to the proposition that spleen morphogenesis may subsequently proceed via one of the following two models.

6.3.2.1 Cell migration: The 'tracking' model

Figure 6.7 illustrates the 'tracking' model. Upon reaching the posterior end of the stomach and forming a connective 'anchor' to the adjacent tissue, the elongating anterior end of the spleen is proposed to "track" along the dorsal mesogastrium. This process is envisaged to be somewhat analagous to limb extension at the apical ectodermal ridge (AER) in that the youngest cells of the developing spleen, having arisen from rapid cell proliferation concentrated towards the anterior domain of the spleno-pancreatic structure, constitute a 'leading edge' which in this case fuses with the adjacent tissue, using the latter as a scaffold. Lagging, developmentally older cells progressively loosen their connections such that the majority of the spleen is a uniform, distinct tissue. Whether the latter separation of cells at the interim occurs purely by physical means or involves cell death is not addressed. By 58 somites, there is no longer a requirement for anchoring structures as the spleen has reached its final position relative to the dorsal stomach and flanks the dorsal mesogastrium.

The 'tracking' model is a means of explaining the observed elongation of the spleen principally in terms of cellular mechanics, in connection with physical properties of the developing foregut, which involves intrinsic factors such as the

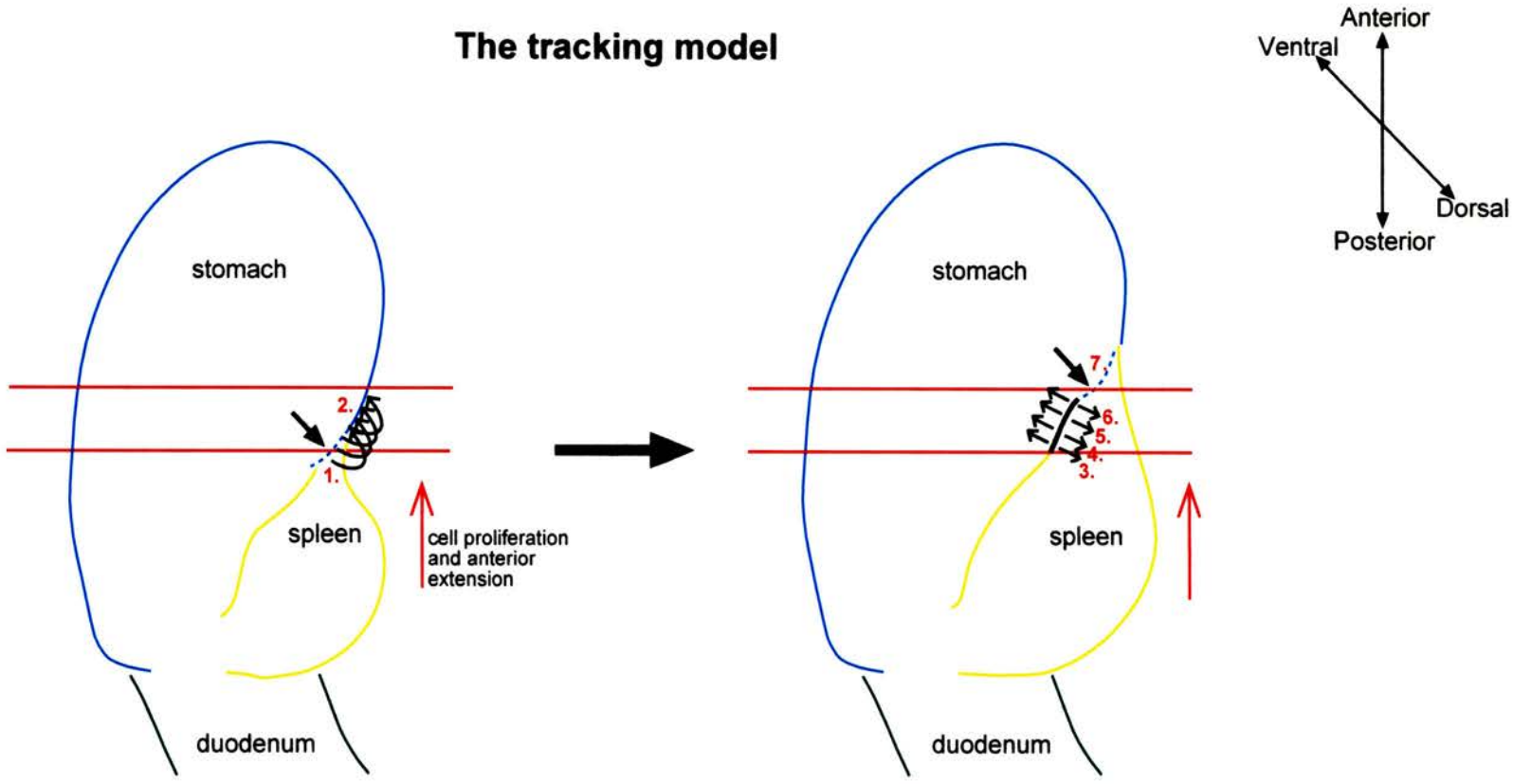


Figure 6.7: A schematic to illustrate the tracking model of spleen morphogenesis. Anterior growth of the spleno-pancreatic tissue towards, and subsequently asymmetrically alongside, the posterior stomach initiates within the 30-40 somite developmental window, by which time the cells at the anterior 'leading' edge of the spleen have formed a connection to the dorsal wall of the developing stomach (1.). Rapid cell proliferation proceeds, fuelling splenogenesis, and the elongating anterior end of the spleen is proposed to "track" along the dorsal mesogastrium. In this process, the youngest cells of the growing spleen, which are envisaged to be those at the anterior 'leading' edge, fuse with the adjacent tissue to form a close connection analogous to a scaffold (2.) which supports and anchors the spleen. Meanwhile, as anterior extension of the spleen domain proceeds, the 'lagging' (developmentally older) cells (originally at position 1.) progressively loosen their connections to the adjacent tissue (3., 4., 5., 6.) such that they become only loosely attached whilst the anterior-most domain of the spleen remains fused (7.). This anterior 'tracking' along the dorsal, righthand wall of the stomach continues until the 55-60 somite stage, by which time the spleen has reached its final position and is attached to the dorsal mesogastrium flanking the anterior stomach. Cell proliferation has presumably slowed resulting in there no longer existing a 'leading edge'.

effect of space constraints on the concomitant growth of organs within the spleno-pancreatic region. Thus growth and extension are initially facilitated by rapid proliferation, which subsequently becomes a self-perpetuating process and migration in an anterior direction constitutes the default developmental pathway.

6.3.2.2 Cell-cell signalling: The 'wave' model

Figure 6.8 illustrates the 'wave' model. In contrast to the 'tracking' model, this proposal invokes a wave of inductive signal beginning at the very posterior end of the stomach to succeed the growth signal provided by the SMP around the 36 somite stage. This novel growth signal would proceed in a "wave-like" manner in a posterior to anterior direction along the dorsal edge of the stomach and marginally ahead of the observed *Hox11* signal, inducing the adjacent proliferation and elongation of the spleen. The position of the anterior spleen 'anchor' along the A-P axis would correspond to the position that the "wave" had reached in the adjacent tissue. Physical separation of cells and/or cell death would occur as in the 'tracking' model to form distinct splenic tissue in regions further from the source of signal owing to the progressive anterior advance of the "wave".

The "wave" hypothesis thus invokes the existence of an as yet unidentified inductive signal for spleen development, expressed in an appropriate manner. If *Hox11* were to constitute this signal, one would expect to detect its expression along the dorsal edge of the stomach in addition to expression in the spleen. Owing to its constituting a simpler explanation for the observed events, and being consistent with the observed *Hox11* expression pattern, I therefore find the 'tracking' model more convincing. Furthermore, it is interesting to note that during a cytological study of the ontogeny of the chick embryonic spleen, rare cells with a distinct ultrastructural morphology indicative of their involvement in a migration process were observed within the mesenchyme at the mouse equivalent of E13.0 (Yassine *et al.*, 1989). These cells thus potentially represent the remnants of a cell population actively involved in anterior migration of the spleen since the developmental stage at which they are noted coincides neatly with the point at which the spleen has adopted its ultimate position with respect to the dorsal wall of the stomach. Further

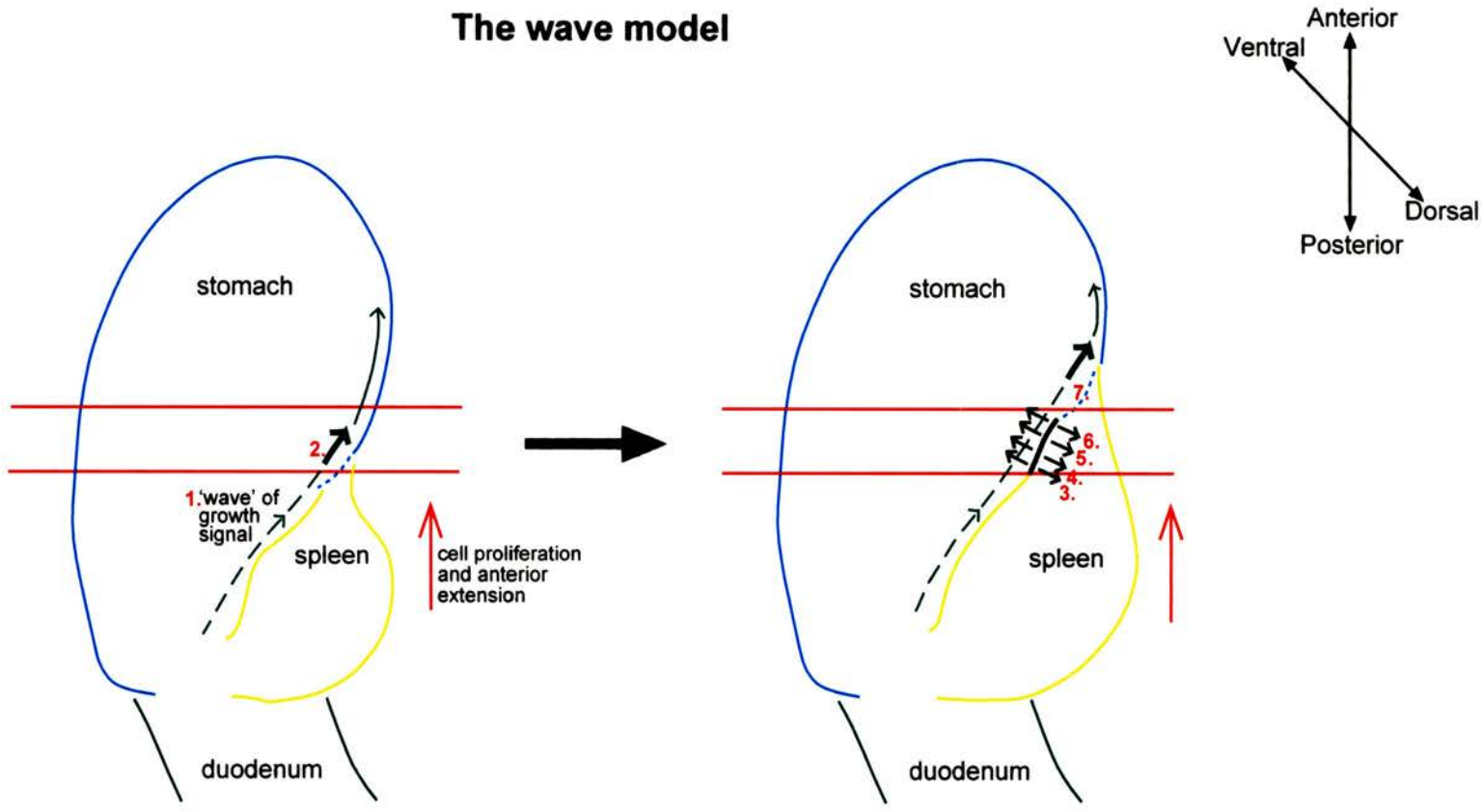


Figure 6.8: A schematic to illustrate the wave model of spleen morphogenesis. In this instance, the anterior growth of the spleno-pancreatic tissue and elongation of the spleen domain is driven by an inductive growth signal which proceeds along the dorsal edge of the stomach in a posterior-to-anterior "wave" of cell-cell signalling (1.), inducing the adjacent proliferation and elongation of the spleen. This growth signal succeeds that potentially provided by the SMP at approximately the 35 somite stage (E10.5) to drive the initial asymmetric growth within the spleno-pancreatic region. The position of the anterior spleen 'anchor' along the A-P axis corresponds to the position that the "wave" of inductive signal has reached in the adjacent tissue (2.). Following passage of the inductive "wave", cells further from the signal progressively lose their tight connections to the adjacent tissue (3., 4., 5., 6.) such that the spleen is only tightly anchored at the anterior-most edge, closest to the source of the inductive signal (7.).

investigations will be required to resolve the present uncertainties and to elucidate the *bona fide* mode of spleen morphogenesis.

6.3.3 Future prospects

OPT undoubtedly possesses enormous potential as a visual aid in developmental studies. It may not currently be possible to visualise tissue in 3D at the ultrastructural level, but for the purpose of following 3D developmental patterning it constitutes both an invaluable resource and a provider of information supplementary to that gleaned by conventional means. Scans performed in glycerol, a necessary prerequisite in this case for preservation of the *Hox11* signal, yield a lower degree of cellular resolution than equivalent samples cleared and scanned in BABB, and due to the nature of the highly viscous glycerol solution, refractive light effects can interfere with gene expression signals. The latter accounts for the non-specific signal occasionally seen within gut tissue. Nevertheless, the present study challenged the limits of OPT and may prove useful in directing future technological developments of this imaging technique.

With regard to the investigation of additional markers expressed in the dorsal pancreatic mesenchyme during the early stages of spleen development, it would be very interesting to co-localise *Islet-1* (*Isl1*) expressing cells and *Hox11* in a 3D context, the one relative to the other. OPT analysis of mutant tissue lacking one of these genes may provide insights into key interactions or cell signalling networks present in the wildtype context upon which spleen development relies. In the case of *Isl1*, this would require the development of a conditional, tissue-specific knockout since *Isl1*-/- mice reportedly die *in utero* around E9.5-E10.0 (Pfaff *et al.*, 1996).

Each of the two models of spleen development discussed above refers to the possibility that cell death potentially accounts for the loosening of the connection between spleen and stomach tissue posterior to the 'anchor'. This could be addressed by TUNEL staining of cross sections of the spleen, in which case apoptosis would be expected to be evident at the junction between the stomach and the spleen. Both the 'tracking' and 'wave' models of spleen morphogenesis also

invoke a higher rate of cell proliferation at the anterior 'leading edge' of the spleen. This could be investigated by BrdU analysis of successive sections through the spleen, or by the development of a 3D BrdU protocol amenable to OPT analysis.

In summary, the present description of spleen morphogenesis establishes a sound foundation for future gene expression studies to further characterise the development of the spleen. Observations made during this study raise fascinating questions as to the nature of the morphogenetic process and form the basis of suggested models for spleen development which can be specifically tested. By extension, such studies contribute to the elucidation of the 3D patterning of the foregut. Importantly, such an approach could be applied to numerous developmental contexts, encompassing a vast evolutionary timescale, and perhaps therefore aid in the establishment of a paradigm for organ development.

Chapter 7 Concluding remarks and future directions

7 Concluding remarks and future directions

This thesis details complementary studies which collectively pertain to an investigation of gut development and the influence of the homeodomain-containing transcription factor *Bapx1* upon this process. *Bapx1* is expressed in a number of mesenchymal domains within the developing mouse embryo and the function of the gene is essential for survival (Lettice *et al.*, 1999; Tribioli and Lufkin, 1999). One of the principal aims at the outset of this research was to determine the role of the mesenchyme in gut development by exploiting the gut-specific expression domain of *Bapx1*. Regulatory sequences controlling the expression of *Bapx1* have yet to be defined, thus a bioinformatics approach was employed to search for putative *Bapx1* regulatory sequences, including control elements responsible for the gut-specific expression. It was envisaged that such elements could subsequently be utilised in mesenchymal cell ablation studies to ablate the *Bapx1*-expressing gut mesenchyme at defined timepoints during embryogenesis. Such studies would permit investigation of both tissue-specific and temporal aspects of the mesenchymal influence upon gut development and perhaps also provide further insight into the role of *Bapx1* in the developing gut.

7.1 ProxB and DistB constitute putative *Bapx1* regulatory loci

The regulatory elements ProxB and DistB, residing 11kb and 18kb downstream of *Bapx1* respectively, constitute promising candidates for loci harbouring *Bapx1* enhancer sequences, as demonstrated by a *LacZ* reporter bioassay. ProxB was able to direct highly tissue-specific expression within the ear/jaw domain during mouse embryogenesis, reflecting the endogenous expression pattern of *Bapx1*. Moreover, conservation of a 180bp portion of mammalian ProxB was discovered at the syntenic locus in zebrafish. A further test of the enhancer capacity of ProxB and its relationship to the endogenous control of *Bapx1* would thus be to replace mammalian ProxB with zebrafish ProxB in the transgenic reporter assay in order to determine whether zebrafish ProxB is able to recapitulate the evolutionarily conserved ear/jaw

domain expression specificities in transgenic mice. Conversely, the reciprocal experiment, introducing the 180bp equivalent region of mammalian ProxB into zebrafish, could also be performed. A *bona fide Bapx1* tissue-specific enhancer would be expected to direct expression to the jaw joint in zebrafish. Such studies would also serve as initial steps in a deletion analysis strategy by defining more precisely the core regulatory sequence responsible for the observed expression pattern. As the amount, quality and availability of avian and amphibian sequence grows, it will be possible to discover whether these organisms also harbour a ProxB sequence, as would be predicted if ProxB does indeed play a part in directing *Bapx1* expression to the ear/jaw.

DistB1 comprises homology blocks 1 and 2 of the more extensive 8kb DistB region which exhibits a high degree of sequence conservation between human and mouse. The discovery that DistB1 is capable of directing tissue-specific reporter gene expression, notably including digit and joint domains resembling those of endogenous *Bapx1*, renders it highly likely that the remaining homology blocks, conserved to an equally high (indeed higher in the case of homology block 6; see Figures 3.5 and 4.1) level, also harbour regulatory sequences. It would thus be of great interest to introduce the remainder of DistB into transgenic mice and to assay for the presence of enhancer sequences. Fluorescence *in situ* hybridisation (FISH) mapping could be employed to identify the chromosomal integration sites of the transgenes and thus explore the relative influence of neighbouring sequences upon the expression pattern exhibited by individual transgenic lines. Several additional CNSs (besides ProxB and DistB) which are conserved between human and mouse and reside both upstream (two CNSs) and downstream (three CNSs) (see Figure 3.4) of *Bapx1* were also noted and it would equally be of interest to include these regions in transgenic reporter assays. Whilst ProxB and DistB1 constitute promising candidates as *Bapx1* enhancers, they do not recapitulate the entire tissue specificity of *Bapx1*. It is thus plausible that the endogenous expression pattern is a manifestation of a *collection* of enhancer sequences clustered around the *Bapx1* locus, and that one or more of the additional conserved regions identified are an integral part of the control cluster.

7.2 Could integral regulatory loci mediate the evolution of the *Bapx1* cluster?

The conservation of synteny in mammals and fish at the *Bapx1* locus between *NM_148894* (upstream of *Bapx1*) and *Rab28* (downstream of *Bapx1*) perhaps implies that important control elements for one or more of these genes are scattered throughout the cluster, thus accounting for the preservation of genomic synteny amongst genes that would appear to have nothing in common besides their genomic location. Breakage of synteny occurs immediately upstream of *NM_148894* in zebrafish and is predicted to be found also in the *Fugu* genome for which, at present, there is insufficient 5' sequence on scaffold_776 (see Figure 3.3). Whilst *Bapx1* exhibits tissue-specific expression, the distribution of EST transcripts of *NM_148894*, *Q9P2L9* and *RAB28* is suggestive of each being a ubiquitously expressed gene. Thus the maintenance of synteny at the *Bapx1* locus potentially results from a scattering of *Bapx1* enhancers across a 270kb mammalian region within which the target gene resides. Notably, such an interpretation can be likened to the situation at the human *DACH* locus whereby the *DACH* gene is bracketed by two extensive gene deserts of 870kb and 1330kb in length. Whilst the genes flanking *DACH* differ in mammals as compared to fish, the linear relationship between conserved non-coding elements is maintained (Nobrega *et al.*, 2003). Few regulatory sequences have been identified in the proximity of the *DACH* promoter (Machon *et al.*, 2002), however comparative sequence analysis and subsequent transgenic bioassays revealed seven elements, sampled from a region greater than 1Mb, that were able to drive *LacZ* in a distinctive set of tissues, recapitulating several aspects of endogenous *DACH* expression (Nobrega *et al.*, 2003). It is notable that the bioinformatics approach adopted in the *DACH* study is comparable to the route followed in the present research and the findings by Nobrega *et al.* support the possibility that a number of *Bapx1* regulatory elements, responsible for various aspects of the endogenous expression pattern, may be scattered within a 270kb syntenic region.

7.2.1 The potential significance of mammal-fish conservation

It is important to reiterate the identification of three separate CNSs conserved to *Fugu* which reside in the fourth intron of *Rab28*, thus towards the 3' end of the *Bapx1* cluster. It is plausible that transgenic reporter assays of these three sequences may reveal enhancer elements responsible for *Bapx1* expression domains common to distant vertebrates as opposed to being mammalian-specific, as in the case of the digits. In particular, non-coding sequences shared between mammals and fish constitute prime candidates in the search for the gut-specific *Bapx1* enhancer. Although DistB1 was demonstrated to direct expression to the gut in three individual transgenic lines, an exact recapitulation of endogenous *Bapx1* expression was not achieved, including an absence of reporter gene transcripts in the spleen, therefore it was concluded that additional elements must exist to refine the observed pattern.

7.3 Could DistB function as a global enhancer?

A further intriguing possibility pertaining to the DistB1-directed expression in several tissues is that the entire DistB CNS may regulate more than one gene within the *Bapx1* cluster. Indeed, the existence of a cluster of global enhancers at the *HoxD* locus that appear to exert an effect on several unrelated genes has recently been reported (Spitz *et al.*, 2003). In the latter study, a transgenic screen of BACs using a transposon-based, locus-targeted enhancer trap methodology revealed a 54kb DNA segment capable of directing reporter gene expression in a profile reminiscent of several structurally, functionally and phylogenetically unrelated genes (*Lunapark*, *Evx2* and *Hoxd* genes) at the *HoxD* locus. Notably, the effect of this so-called global control region (GCR) spans 250kb. The GCR was further refined, by comparative sequence analysis, to 40kb in mammals for which an 8kb *Fugu* equivalent was discovered. The pufferfish GCR did not appear capable of directing expression to the digits, a tetrapod-specific domain. Interestingly, whilst DistB1 directed expression in a tissue-specific manner to several regions of endogenous *Bapx1* expression, for example the digits and limb joints, reporter gene transcripts were also detected in several tissues not associated with *Bapx1* function. A number of these

additional domains correspond to regions for which *NM_148894*-, *Q9P2L9*- and *RAB28*-derived EST transcripts have been reported. It is conceivable that DistB enhancer sequences may contribute to the control of, and thus be shared by, the latter three genes, particularly if these seemingly ubiquitous genes possess low promoter specificities.

7.4 Alternative strategies to investigate the role of the mesenchyme

Considering the success of employing the Tn7 transposition system (Biery *et al.*, 2000) in the *HoxD* study by Spitz *et al.*, it is feasible that this strategy could be modified to replace certain stages of the recombineering approach designed to achieve mesenchymal cell ablation via *Bapx1*-directed tissue-specific expression of a cytotoxic gene (detailed in Chapter 3). The Tn7 approach was previously attempted in our lab without success, however integral properties of the particular genomic target site may be important determinants of the efficacy of the system, as also proposed for modification by recombineering. Thus the Tn7 β-*lac* reporter transposon could initially be used to tag *Bapx1*-containing PACs or BACs by *in vitro* transposition. Subsequent identification would be required of vectors containing the Tn7 transposon inserted into *Bapx1* coding sequence and thus deemed to fall under the control of endogenous *Bapx1* regulators. Notably, Spitz *et al.* report targeting efficiencies of between 0.8% and 6%, routine achievement of a unique insertion event, and randomly distributed insertion sites such that in a pilot experiment involving a 175kb BAC, insertions were recovered nearly every 10kb. The reporter transposon, having been assayed in transgenic mice to ensure appropriate *Bapx1* expression, would then be modified by the replacement of the β-globin promoter/*LacZ* reporter gene in the Tn7 transposon with a cytotoxic gene, for example the diptheria toxin receptor. This latter stage could potentially be achieved by recombineering, although a modification to the protocol formerly adopted would be required. The design of longer homology arms on targeting cassettes would perhaps overcome the limited success achieved with 50nt homology arms on

cassettes targeted to the *Bapx1* locus in PACs. The benefit of including more extensive regions of flanking homology in a recombineering protocol has recently been demonstrated in the generation of conditional knockout mice (Liu *et al.*, 2003).

Embryonic stem (ES) cell modification constitutes a further means of modifying the *Bapx1* locus, however as mentioned in Chapter 3, this procedure is highly time-consuming, expensive and technically difficult. A different approach altogether to effect ablation of *Bapx1*-expressing mesenchymal cells in transgenic mice would be to create a transgenic construct in which the cytotoxic gene is engineered downstream of a *Bapx1* tissue-specific enhancer. The ability to use such a strategy is dependent upon the results of the additional transgenic bioassays proposed above. If one or more of the candidate CNSs within the *Bapx1* cluster exhibits tissue-specific expression, in particular if a gut-specific enhancer was to be identified, these sequences could be cloned directly upstream of the cytotoxic gene, thus bypassing the requirement for complex and time-consuming experiments to modify the endogenous *Bapx1* locus.

7.5 Investigating the position of *Bapx1* in the signalling pathway underlying gut development

To achieve a greater understanding of the role of *Bapx1* in gut organogenesis will require detailed investigations at the functional genomics level, including thorough exploration of the interaction of Bapx1 with other signalling molecules. The assignment of a number of Bapx1 interactors to a common molecular cascade controlling embryonic gut development will represent a significant advancement in the elucidation of the intricacies of gut organogenesis.

7.5.1 Screening for downstream targets of *Bapx1*

Microarrays constitute a high throughput technique whereby it is possible to gain a more comprehensive picture of the simultaneously occurring interactions amongst thousands of genes. Comparisons of gene expression levels would be made between

microarrays screened with wildtype and *Bapx1* mutant tissue to identify candidate genes affected by the loss of *Bapx1*. Whilst numerous false positives may be identified owing to variant cellular conditions at a given time, there is equally the potential to identify candidate downstream targets of *Bapx1* in a signalling cascade.

7.5.2 Putative DNA-protein interactions

In order to identify genomic sequences with which Bapx1 interacts it is advantageous to have first established a number of candidate molecules. To this effect, microarray technology can be employed to generate a number of potential interactors. Similarly, to establish the identity of factors acting upstream of *Bapx1*, which may bind to sequences at the *Bapx1* locus, benefits from the existence of defined candidates. With regard to the latter, genetic analyses of *Pax1* and *Pax9* mutant mice provided a good indication that they may interact directly with the *Bapx1* locus in the control of chondrogenesis (Rodrigo *et al.*, 2003). Subsequent *in vitro* binding assays demonstrated that Pax1 and Pax9 physically interact with the *Bapx1* promoter.

7.5.3 Putative protein-protein interactions

As in the case of DNA-protein interactions, the elucidation of potential protein-protein interactions involving Bapx1 would benefit from the prior establishment of signalling molecules which reside in the same molecular cascade as Bapx1. This could be accomplished, for example, by the use of microarray screens (to identify candidate downstream factors) or through genetic studies of knockout mice in which a similar pathway is affected by loss of the given gene. Yeast two-hybrid technology could subsequently be employed to identify novel protein-protein interactions which could further be verified *in vivo* by biochemical means.

7.5.4 Putative DNA-protein interactions involving *Bapx1* regulatory elements

In addition to assays involving the Bapx1 protein, studies involving regulatory elements of the gene could also provide vital clues as to its position within a molecular signalling network. Thus candidate regulatory elements, such as ProxB and DistB1, would form DNA sequences to which certain proteins may have the ability to bind. In the case of ProxB and DistB1 (in addition to other candidate CNSs residing at the *Bapx1* locus), the efficacy of employing a standard biochemical approach such as an electrophoretic mobility shift assay (EMSA) would be limited by the absence of a known or candidate interacting protein. In turn, a lack of suitable biological material from which to extract sufficient cellular protein would probably be a complicating factor. Nevertheless, the screening of a lambda (λ) cDNA expression library with a short, radiolabelled region within a putative regulatory element constitutes one means of identifying proteins which may bind to the regulatory sequence *in vivo*.

Alternatively, short DNA target "bait" sequences derived from putative regulatory elements could be cloned upstream of a yeast reporter gene and DNA-protein interactions subsequently screened in a yeast one-hybrid system. Notably, yeast One-Hybrid Library Construction and Screening Kits (Clontech; BD Biosciences) have recently been developed which facilitate the rapid construction and screening of a cDNA library of choice. Thus a 'personalised' cDNA expression library, for example derived from branchial arch tissue (to investigate potential ProxB interactions) or the spleno-pancreatic region, can be constructed as dictated by the investigator.

Any proteins found to interact with the conserved sequences, some of which have been demonstrated herein to possess enhancer capacities and potentially function in the endogenous control of *Bapx1* (ProxB and DistB1), would constitute candidate mediators of the endogenous control of *Bapx1*. An additional level of control at the *Bapx1* locus is perhaps provided by self-regulation of *Bapx1* expression via the Bapx1 peptide. Such auto-regulation could potentially be achieved by the binding of Bapx1 to enhancer sequences at the *Bapx1* locus.

Accordingly, *in vitro* binding assays, for example EMSA, could be performed with a selection of CNSs within the *Bapx1* cluster, together with the Bapx1 protein.

7.5.5 Implications for further characterisation of molecular boundaries within the gut

The discovery of additional proteins implicated in the *Bapx1* signalling cascade would provide further markers to study which potentially play a role, not only in terms of gut development *per se*, but also more specifically in the establishment of distinct molecular boundaries within the foregut. The study of defined expression domains and their subsequent disruption in the *Bapx1* mutant would be further extended by investigation of various genes, together with enzymes such as alkaline phosphatase, that exhibit specific expression patterns within the gut. The identification and characterisation of such markers will likely rely on the integration and interpretation of experimental results derived from a whole host of individual studies conducted using a variety of evolutionarily diverged organisms. The extent to which the function of *Bapx1* in gut development is evolutionarily conserved will be an important determinant in the ability to directly apply preliminary results in one organism to investigations in those having undergone extensive divergence. As an example, upregulation of *Bmp4* expression was not detected in the posterior stomach of mice in the absence of *Bapx1* (Chapter 5), despite expectations of altered transcript distribution based on analogous studies conducted in chick. Thus multi-species studies have the potential to impart knowledge regarding evolutionary changes (structural, functional and/or in terms of regulatory sequences) that a particular locus may have undergone. In this respect, an alteration to the regulatory capacities of *Bapx1* regarding downstream effectors in craniofacial development has recently been demonstrated (Tucker *et al.*, 2003, submitted; see Introduction).

Hox genes are important in the patterning of the embryo along the A-P axis and certain members are reportedly expressed in defined domains within the gut of both invertebrates (LeMotte *et al.*, 1989; Tremml and Bienz, 1989; Reuter and Scott, 1990) and vertebrates (Roberts *et al.*, 1995; Yokouchi *et al.*, 1995; Sakiyama *et al.*,

2001; Kawazoe *et al.*, 2002). Interestingly, *Hoxa5* has been implicated in the control of molecules involved in epithelial-mesenchymal signalling within the mouse stomach. In the absence of *Hoxa5* function, the expression of *Shh*, *Ihh*, TGFβ family members such as *Bmp4*, and *Fgf10* is altered (Aubin *et al.*, 2002). A recent expression profile analysis of multiple *Wnt* genes in the chick gut (Theodosiou and Tabin, 2003) also provides a battery of genes amongst which there may be important *Bapx1* interactors whose functions are critical to the normal patterning of the gut. It may therefore be beneficial to perform extensive expression analyses of the *Hox*, *Wnt* and, considering their potential involvement in spleno-pancreatic outgrowth, *Fgf* gene families. Further analysis of the gross malformation manifest in the *Bapx1* mutant gut is imperative to our understanding of the molecular mechanisms responsible for the posterior stomach expansion, duodenal fusion and abnormal branching phenotypes. Current investigations involve the use of multiple pancreatic and endodermal markers in order to determine the altered locations of the spleno-pancreatic organs and the cellular identities of mutant structures unique to *Bapx1-/-* animals. As is generally the rationale behind studies of null mutant organisms, detailed analysis of the *Bapx1-/-* gut phenotype is expected to provide considerable insights with respect to the endogenous role of *Bapx1* in gut development.

7.6 Morphogenesis of the spleen

One of the phenotypic manifestations of the loss of *Bapx1* to which little attention has thus far been paid is asplenia. The function of *Bapx1* in the spleen is presently unknown although it presumably resides upstream of *Nkx2.5* and *Hox11* since splenic expression of the latter two genes is lost in the absence of *Bapx1* (Lettice *et al.*, 1999; Hecksher-Sørensen *et al.*, 2003, submitted). *Hox11* has been demonstrated to act cell autonomously to control the differentiation of splenic precursor cells (Kanzler and Dear, 2001). Prior to assigning a specific role for any of the aforementioned molecules in splenogenesis, it would be advantageous to possess more detailed knowledge regarding the morphogenesis of the spleen, a subject for which there is a paucity of data. Chapter 6 addresses the morphological development

of the spleen and two alternative models are proposed to explain the observed elongation of the splenic domain within a limited developmental window. The resolution of these models could potentially be aided by ultrastructural studies involving the use of electron microscopy. Spleen extension via the "tracking" hypothesis invokes the existence of cells with differing properties to those induced to differentiate into spleen by an adjacent "wave"-like signal. The "tracking" model suggests that the adhesive power of cells is an important determinant in their ability to use the adjacent tissue as a scaffold to direct growth. Accordingly it may be possible to detect, at the ultrastructural level, cellular processes such as filipodia at the stomach-spleen junction that would aid such migration of splenic cells.

Apoptosis would perhaps be predicted to play a greater role in the "wave" model than in the "tracking" model, to effect physical separation of stomach and spleen cells posterior to the 'leading edge' of the spleen and thus further from the hypothesised source of inductive signal. Whilst cell death may also play a role in the "tracking" model, it is not strictly required to effect the separation of tissue at the stomach-spleen junction. The principal point of contact by "tracking" occurs at the anterior edge and movement may rely instead on structural changes to those cells at the 'leading edge' to allow tight adherence to the adjacent scaffold. Thus TUNEL staining may provide supporting evidence for one or other of the proposed models.

Establishment of an organ culture system in which to follow gut development potentially constitutes a further means of distinguishing between the two hypothesised models of spleen development. A growth matrix would be required upon which the 3D conformation of the foregut, particularly the 'scaffolding' stomach, is maintained. The injection of a fluorescent dye which does not kill cells, such as diI, would effect the labelling of a specific cell population. Such a culture system could be used to study spleen development at the critical stage between E10.5 and E11.0 at which time the rapid elongation process is evidently initiated. A preliminary approach would be to fluorescently label cells at regular intervals along the dorsal edge of the stomach and follow development for up to 12 hours. If labelled cells were subsequently incorporated into the growing spleen, this would indicate that spleen development proceeds via the "wave"-like model. Conversely, if the elongating spleen by-passed the labelled cells, which instead remained within a

separate structure (the stomach), support for the "tracking" model of splenogenesis is provided. A gut organ culture system would allow for a number of additional experiments to probe various aspects of gut development both in wildtype organisms and mutants, at specific developmental stages and using various gut markers. The manipulation and recombination of mutant and wildtype gut tissue could provide fascinating insights into the genetic, biochemical and cellular conditions required for the wildtype developmental programme to proceed.

7.6.1 To what extent does the spleen determine the ultimate location of the pancreas?

Considering the altered morphology of the spleen and pancreas in the absence of functional *Shh*, it may prove beneficial to follow a time course of spleen development in *Shh* mutant gut tissue comparable to that performed on wildtype tissue. Analysis of the morphological development of the mutant spleno-pancreatic region may aid in the establishment of a model for splenogenesis. Chapter 6 illustrates the early movement of the spleno-pancreatic structure as a mass of tissue whereby *Hox11* (marking splenic precursor cells) is expressed in the anterior domain of the tissue at the 43 somite stage. Thus an intriguing question to be addressed is the relative importance of the spleen in determining the ultimate position of the pancreas. Having established a model for the morphological development of the spleen, it will be most interesting to analyse developmental progress in the asplenic *Bapx1*-/- mice for which the ultimate asymmetric positioning of the pancreas adjacent to the stomach appears to be maintained. In the absence of a spleen therefore, what guides the pancreas and determines its final positioning, and how does morphological development compare to the wildtype situation? OPT analysis will likely provide useful insights into the mutant developmental process by facilitating studies of asplenic mutants in a 3D context.

7.7 Closing statement

In summary, the work presented herein relates to fundamental aspects of embryonic development: the influence of regulatory sequences, the establishment of specific molecular domains and the importance of morphogenetic movements in the determination of the definitive body plan. That gut organogenesis alone represents one of numerous processes occurring simultaneously during development alludes to both the enormity and extraordinary complexity of the evolutionary task to generate a working protocol for mammalian development.

References

Ahlgren,U., Jonsson,J., & Edlund,H. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development* **122**, 1409-1416 (1996).

Ahlgren,U., Pfaff,S.L., Jessell,T.M., Edlund,T., & Edlund,H. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature* **385**, 257-260 (1997).

Akazawa,H., Komuro,I., Sugitani,Y., Yazaki,Y., Nagai,R., & Noda,T. Targeted disruption of the homeobox transcription factor Bapx1 results in lethal skeletal dysplasia with asplenia and gastroduodenal malformation. *Genes Cells* **5**, 499-513 (2000).

Aparicio,S., Morrison,A., Gould,A., Gilthorpe,J., Chaudhuri,C., Rigby,P., Krumlauf,R., & Brenner,S. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. *Proc. Natl. Acad. Sci. U. S. A* **92**, 1684-1688 (1995).

Aparicio,S., Chapman,J., Stupka,E., Putnam,N., Chia,J.M., Dehal,P., Christoffels,A., Rash,S., Hoon,S., Smit,A., Gelpke,M.D., Roach,J., Oh,T., Ho,I.Y., Wong,M., Detter,C., Verhoef,F., Predki,P., Tay,A., Lucas,S., Richardson,P., Smith,S.F., Clark,M.S., Edwards,Y.J., Doggett,N., Zharkikh,A., Tavtigian,S.V., Pruss,D., Barnstead,M., Evans,C., Baden,H., Powell,J., Glusman,G., Rowen,L., Hood,L., Tan,Y.H., Elgar,G., Hawkins,T., Venkatesh,B., Rokhsar,D., & Brenner,S. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* **297**, 1301-1310 (2002).

Apelqvist,A., Ahlgren,U., & Edlund,H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr. Biol.* **7**, 801-804 (1997).

Aubin,J., Dery,U., Lemieux,M., Chailler,P., & Jeannotte,L. Stomach regional specification requires Hoxa5-driven mesenchymal-epithelial signaling. *Development* **129**, 4075-4087 (2002).

Avner,P. & Heard,E. X-chromosome inactivation: counting, choice and initiation. *Nat. Rev. Genet.* **2**, 59-67 (2001).

Azpiazu,N. & Frasch,M. tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. *Genes Dev.* **7**, 1325-1340 (1993).

Bard,J.B.L. Morphogenesis. (Cambridge University Press, 1990).

Bellusci,S., Henderson,R., Winnier,G., Oikawa,T., & Hogan,B.L. Evidence from normal expression and targeted misexpression that bone morphogenetic protein (Bmp-4) plays a role in mouse embryonic lung morphogenesis. *Development* **122**, 1693-1702 (1996).

Ben Yair,R., Kahane,N., & Kalcheim,C. Coherent development of dermomyotome and dermis from the entire mediolateral extent of the dorsal somite. *Development* **130**, 4325-4336 (2003).

Bhushan,A., Itoh,N., Kato,S., Thiery,J.P., Czernichow,P., Bellusci,S., & Scharfmann,R. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development* **128**, 5109-5117 (2001).

Biery,M.C., Stewart,F.J., Stellwagen,A.E., Raleigh,E.A., & Craig,N.L. A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res.* **28**, 1067-1077 (2000).

Bitgood,M.J. & McMahon,A.P. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126-138 (1995).

Borsani,G., Tonlorenzi,R., Simmler,M.C., Dandolo,L., Arnaud,D., Capra,V., Grompe,M., Pizzuti,A., Muzny,D., Lawrence,C., & . Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-329 (1991).

Branford,W.W., Essner,J.J., & Yost,H.J. Regulation of gut and heart left-right asymmetry by context-dependent interactions between xenopus lefty and BMP4 signaling. *Dev. Biol.* **223**, 291-306 (2000).

Brauers,A., Schurmann,A., Massmann,S., Muhl-Zurbes,P., Becker,W., Kainulainen,H., Lie,C., & Joost,H.G. Alternative mRNA splicing of the novel GTPase Rab28 generates isoforms with different C-termini. *Eur. J. Biochem.* **237**, 833-840 (1996).

Brenner,S., Elgar,G., Sandford,R., Macrae,A., Venkatesh,B., & Aparicio,S. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature* **366**, 265-268 (1993).

Brockdorff,N., Ashworth,A., Kay,G.F., Cooper,P., Smith,S., McCabe,V.M., Norris,D.P., Penny,G.D., Patel,D., & Rastan,S. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329-331 (1991).

Brown,C.J., Ballabio,A., Rupert,J.L., Lafreniere,R.G., Grompe,M., Tonlorenzi,R., & Willard,H.F. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38-44 (1991).

Brown,C.J., Hendrich,B.D., Rupert,J.L., Lafreniere,R.G., Xing,Y., Lawrence,J., & Willard,H.F. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).

Bulyk,M.L., Huang,X., Choo,Y., & Church,G.M. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A* **98**, 7158-7163 (2001).

Butler,H. & Juurlink,B.H.J. An atlas for staging mammalian and chick embryos. (CRC Press Inc., Florida, 1987).

Carlson,B.M. Human Embryology and Developmental Biology. (Mosby Yearbook, St. Louis, MO., 1999).

Carroll,S.B., Grenier,J.K., & Weatherbee,S.D. From DNA to diversity: molecular genetics and the evolution of animal design. (Blackwell Science, Malden, MA., 2001).

Chang,C., Wilson,P.A., Mathews,L.S., & Hemmati-Brivanlou,A. A Xenopus type I activin receptor mediates mesodermal but not neural specification during embryogenesis. *Development* **124**, 827-837 (1997).

Choi,O.R. & Engel,J.D. A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult beta-globin gene. *Nature* **323**, 731-734 (1986).

Copeland,N.G., Jenkins,N.A., & Court,D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769-779 (2001).

Davies,J.A. Morphogenesis of the metanephric kidney. *Scientific World Journal.* **2**, 1937-1950 (2002).

Dear,T.N., Colledge,W.H., Carlton,M.B., Lavenir,I., Larson,T., Smith,A.J., Warren,A.J., Evans,M.J., Sofroniew,M.V., & Rabbitts,T.H. The Hox11 gene is essential for cell survival during spleen development. *Development* **121**, 2909-2915 (1995).

Deutsch,U., Dressler,G.R., & Gruss,P. Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53**, 617-625 (1988).

Durand,D. Vertebrate evolution: doubling and shuffling with a full deck. *Trends Genet.* **19**, 2-5 (2003).

Echelard,Y., Epstein,D.J., St Jacques,B., Shen,L., Mohler,J., McMahon,J.A., & McMahon,A.P. Sonic hedgehog, a member of a family of putative signalling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417-1430 (1993).

Ellis,H.M., Yu,D., DiTizio,T., & Court,D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A* **98**, 6742-6746 (2001).

Erdmann,V.A., Szymanski,M., Hochberg,A., Groot,N., & Barciszewski,J. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res.* **28**, 197-200 (2000).

Erdmann,V.A., Barciszewska,M.Z., Hochberg,A., de Groot,N., & Barciszewski,J. Regulatory RNAs. *Cell Mol. Life Sci.* **58**, 960-977 (2001a).

Erdmann,V.A., Barciszewska,M.Z., Szymanski,M., Hochberg,A., de Groot,N., & Barciszewski,J. The non-coding RNAs as riboregulators. *Nucleic Acids Res.* **29**, 189-193 (2001b).

Erlebacher,A., Filvaroff,E.H., Gitelman,S.E., & Derynck,R. Toward a molecular understanding of skeletal development. *Cell* **80**, 371-378 (1995).

Esni,F., Johansson,B.R., Radice,G.L., & Semb,H. Dorsal pancreas agenesis in N-cadherin-deficient mice. *Dev. Biol.* **238**, 202-212 (2001).

Florea,L., Hartzell,G., Zhang,Z., Rubin,G.M., & Miller,W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**, 967-974 (1998).

Fukuda,K. & Yasugi,S. Versatile roles for sonic hedgehog in gut development. *J. Gastroenterol.* **37**, 239-246 (2002).

Fukuda,K., Kameda,T., Saitoh,K., Iba,H., & Yasugi,S. Down-regulation of endodermal Shh is required for gland formation in chicken stomach. *Mech. Dev.* **120**, 801-809 (2003).

Gans,C. & Northcutt,R.G. Neural Crest and the Origin of Vertebrates: A New Head. *Science* **220**, 268-274 (1983).

Gilbert,S.F. Developmental Biology. (Sinauer Associates, Inc., 1997).

Golosow,N. & Grobstein,C. Epitheliomesenchymal interaction in pancreatic morphogenesis. *Dev. Biol.* **4**, 242-255 (1962).

Green,M.C. A defect of the splanchnic mesoderm caused by the mutant gene dominant hemimelia in the mouse. *Dev. Biol.* **15**, 62-89 (1967).

Greene,W.K., Bahn,S., Masson,N., & Rabbitts,T.H. The T-cell oncogenic protein HOX11 activates Aldh1 expression in NIH 3T3 cells but represses its expression in mouse spleen development. *Mol. Cell Biol.* **18**, 7030-7037 (1998).

Gruenwald,P. Normal and abnormal detachment of body and gut from the blastoderm in the chick embryo, with remarks on the early development of the allantois. *J. Morphol.* **69**, 83-125 (1941).

Hamada,H., Meno,C., Watanabe,D., & Saijoh,Y. Establishment of vertebrate left-right asymmetry. *Nat. Rev. Genet.* **3**, 103-113 (2002).

Hamburger,V. & Hamilton,H.L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49-92 (1951).

Hammerschmidt,M., Brook,A., & McMahon,A.P. The world according to hedgehog. *Trends Genet.* **13**, 14-21 (1997).

Hardison,R.C., Oeltjen,J., & Miller,W. Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* **7**, 959-966 (1997).

Hardison,R.C. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**, 369-372 (2000).

Harrison,K.A., Thaler,J., Pfaff,S.L., Gu,H., & Kehrl,J.H. Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. *Nat. Genet.* **23**, 71-75 (1999).

Harvey,R.P. NK-2 homeobox genes and heart development. *Dev. Biol.* **178**, 203-216 (1996).

Have-Opbroek,A.A. Lung development in the mouse embryo. *Exp. Lung Res.* **17**, 111-130 (1991).

Hebrok,M., Kim,S.K., & Melton,D.A. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes Dev.* **12**, 1705-1713 (1998).

Hebrok,M., Kim,S.K., St Jacques,B., McMahon,A.P., & Melton,D.A. Regulation of pancreas development by hedgehog signalling. *Development* **127**, 4905-4913 (2000).

Hebrok,M. Hedgehog signalling in pancreas development. *Mech. Dev.* **120**, 45-57 (2003).

Hecksher-Sørensen,J. A developmental analysis of the mouse mutant Dominant hemimelia (*Dh*). (Thesis, 2001).

Hedges,S.B. & Kumar,S. Genomics. Vertebrate genomes compared. *Science* **297**, 1283-1285 (2002).

Herbrand,H., Pabst,O., Hill,R., & Arnold,H.H. Transcription factors Nkx3.1 and Nkx3.2 (Bapx1) play an overlapping role in sclerotomal development of the mouse. *Mech. Dev.* **117**, 217-224 (2002).

Herzer,U., Crocoll,A., Barton,D., Howells,N., & Englert,C. The Wilms tumor suppressor gene wt1 is required for development of the spleen. *Curr. Biol.* **9**, 837-840 (1999).

Higgins,D.G., Thompson,J.D., & Gibson,T.J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383-402 (1996).

Hoch,M. & Pankratz,M.J. Control of gut development by fork head and cell signalling molecules in Drosophila. *Mech. Dev.* **58**, 3-14 (1996).

Holland,P.W., Garcia-Fernandez,J., Williams,N.A., & Sidow,A. Gene duplications and the origins of vertebrate development. *Dev. Suppl* 125-133 (1994).

Hourdry,J., L'Hermite,A., & Ferrand,R. Changes in the digestive system and feeding behaviour of anuran amphibians during metamorphosis. *Physiological and Biochemical Zoology* **69**, 219-251 (1996).

Hurst,L.D. & Smith,N.G. Molecular evolutionary evidence that H19 mRNA is functional. *Trends Genet.* **15**, 134-135 (1999).

Ingham,P.W. & Fietz,M.J. Quantitative effects of hedgehog and decapentaplegic activity on the patterning of the Drosophila wing. *Curr. Biol.* **5**, 432-440 (1995).

Johansson,O., Alkema,W., Wasserman,W.W., & Lagergren,J. Identification of functional clusters of transcription factor binding motifs in genome sequences: the MSCAN algorithm. *Bioinformatics.* **19 Suppl 1**, I169-I176 (2003).

Jonsson,J., Carlsson,L., Edlund,T., & Edlund,H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606-609 (1994).

Kanzler,B. & Dear,T.N. Hox11 acts cell autonomously in spleen development and its absence results in altered cell fate of mesenchymal spleen precursors. *Dev. Biol.* **234**, 231-243 (2001).

Kaufman,M.H. The Atlas of Mouse Development. (Academic Press Ltd., London, 1992).

Kawazoe,Y., Sekimoto,T., Araki,M., Takagi,K., Araki,K., & Yamamura,K. Region-specific gastrointestinal Hox code during murine embryonal gut development. *Dev. Growth Differ.* **44**, 77-84 (2002).

Kedinger,M., Simon-Assmann,P.M., Lacroix,B., Marxer,A., Hauri,H.P., & Haffen,K. Fetal gut mesenchyme induces differentiation of cultured intestinal endodermal and crypt cells. *Dev. Biol.* **113**, 474-483 (1986).

Kedinger,M., Simon-Assmann,P., Bouziges,F., & Haffen,K. Epithelial-mesenchymal interactions in intestinal epithelial differentiation. *Scand. J. Gastroenterol. Suppl* **151**, 62-69 (1988).

Kedinger,M., Simon-Assmann,P., Bouziges,F., Arnold,C., Alexandre,E., & Haffen,K. Smooth muscle actin expression during rat gut development and induction in fetal skin fibroblastic cells associated with intestinal embryonic epithelium. *Differentiation* **43**, 87-97 (1990).

Kim,S.K., Hebrok,M., & Melton,D.A. Notochord to endoderm signaling is required for pancreas development. *Development* **124**, 4243-4252 (1997).

Kim,S.K., Hebrok,M., Li,E., Oh,S.P., Schrewe,H., Harmon,E.B., Lee,J.S., & Melton,D.A. Activin receptor patterning of foregut organogenesis. *Genes Dev.* **14**, 1866-1871 (2000).

Koehler,K., Franz,T., & Dear,T.N. Hox11 is required to maintain normal Wt1 mRNA levels in the developing spleen. *Dev. Dyn.* **218**, 201-206 (2000).

Kohler,J., Schafer-Preuss,S., & Buttgereit,D. Related enhancers in the intron of the beta1 tubulin gene of Drosophila melanogaster are essential for maternal and CNS-specific expression during embryogenesis. *Nucleic Acids Res.* **24**, 2543-2550 (1996).

Krauss,S., Concordet,J.P., & Ingham,P.W. A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* **75**, 1431-1444 (1993).

Kreidberg,J.A., Sariola,H., Loring,J.M., Maeda,M., Pelletier,J., Housman,D., & Jaenisch,R. WT-1 is required for early kidney development. *Cell* **74**, 679-691 (1993).

Kumar,M., Jordan,N., Melton,D., & Grapin-Botton,A. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev. Biol.* **259**, 109-122 (2003).

Kumar,S., Tamura,K., Jakobsen,I.B., & Nei,M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics.* **17**, 1244-1245 (2001).

Lammert,E., Cleaver,O., & Melton,D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564-567 (2001).

Lee,E.C., Yu,D., Martinez de,V., Tessarollo,L., Swing,D.A., Court,D.L., Jenkins,N.A., & Copeland,N.G. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65 (2001).

Lee,K.J., Dietrich,P., & Jessell,T.M. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. *Nature* **403**, 734-740 (2000).

LeMotte,P.K., Kuroiwa,A., Fessler,L.I., & Gehring,W.J. The homeotic gene Sex Combs Reduced of Drosophila: gene structure and embryonic expression. *EMBO J.* **8**, 219-227 (1989).

Lenhard,B., Sandelin,A., Mendoza,L., Engstrom,P., Jareborg,N., & Wasserman,W.W. Identification of conserved regulatory elements by comparative genome analysis. *J. Biol.* **2**, 13 (2003).

Lettice,L., Hecksher-Sorensen,J., & Hill,R. The role of Bapx1 (Nkx3.2) in the development and evolution of the axial skeleton. *J. Anat.* **199**, 181-187 (2001).

Lettice,L.A., Purdie,L.A., Carlson,G.J., Kilanowski,F., Dorin,J., & Hill,R.E. The mouse bagpipe gene controls development of axial skeleton, skull, and spleen. *Proc. Natl. Acad. Sci. U. S. A* **96**, 9695-9700 (1999).

Lettice,L.A., Heaney,S.J., Purdie,L.A., Li,L., de Beer,P., Oostra,B.A., Goode,D., Elgar,G., Hill,R.E., & de Graaff,E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725-1735 (2003).

Lewin,B. Genes VII. (Oxford University Press, 2000).

Li,H., Arber,S., Jessell,T.M., & Edlund,H. Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. *Nat. Genet.* **23**, 67-70 (1999).

Liu,P., Jenkins,N.A., & Copeland,N.G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476-484 (2003).

Logan,C., Wingate,R.J., McKay,I.J., & Lumsden,A. Tlx-1 and Tlx-3 homeobox gene expression in cranial sensory ganglia and hindbrain of the chick embryo: markers of patterned connectivity. *J. Neurosci.* **18**, 5389-5402 (1998).

Loots,G.G., Locksley,R.M., Blankespoor,C.M., Wang,Z.E., Miller,W., Rubin,E.M., & Frazer,K.A. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136-140 (2000).

Lu,J., Chang,P., Richardson,J.A., Gan,L., Weiler,H., & Olson,E.N. The basic helix-loop-helix transcription factor capsulin controls spleen organogenesis. *Proc. Natl. Acad. Sci. U. S. A* **97**, 9525-9530 (2000).

Lyons,I., Parsons,L.M., Hartley,L., Li,R., Andrews,J.E., Robb,L., & Harvey,R.P. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeobox gene Nkx2-5. *Genes Dev.* **9**, 1654-1666 (1995).

Lyons,K.M., Pelton,R.W., & Hogan,B.L. Organogenesis and pattern formation in the mouse: RNA distribution patterns suggest a role for bone morphogenetic protein-2A (BMP-2A). *Development* **109**, 833-844 (1990).

Lyons,K.M., Hogan,B.L., & Robertson,E.J. Colocalization of BMP 7 and BMP 2 RNAs suggests that these factors cooperatively mediate tissue interactions during murine development. *Mech. Dev.* **50**, 71-83 (1995).

Machon,O., van den Bout,C.J., Backman,M., Rosok,O., Caubit,X., Fromm,S.H., Geronimo,B., & Krauss,S. Forebrain-specific promoter/enhancer D6 derived from the mouse Dach1 gene controls expression in neural stem cells. *Neuroscience* **112**, 951-966 (2002).

MacKenzie,A., Purdie,L., Davidson,D., Collinson,M., & Hill,R.E. Two enhancer domains control early aspects of the complex expression pattern of Msx1. *Mech. Dev.* **62**, 29-40 (1997).

Mallo,M. Retinoic acid disturbs mouse middle ear development in a stage-dependent fashion. *Dev. Biol.* **184**, 175-186 (1997).

Manning,M.J. & Horton,J.D. Histogenesis of lymphoid organs in larvae of the South African clawed toad, Xenopus laevis (Daudin). *J. Embryol. Exp. Morphol.* **22**, 265-277 (1969).

Mayor,C., Brudno,M., Schwartz,J.R., Poliakov,A., Rubin,E.M., Frazer,K.A., Pachter,L.S., & Dubchak,I. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics.* **16**, 1046-1047 (2000).

McPherron,A.C., Lawler,A.M., & Lee,S.J. Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat. Genet.* **22**, 260-264 (1999).

Miletich,I. & Sharpe,P.T. Normal and abnormal dental development. *Hum. Mol. Genet.* **12**, R69-R73 (2003).

Miller,C.T., Yelon,D., Stainier,D.Y., & Kimmel,C.B. Two endothelin 1 effectors, hand2 and bapx1, pattern ventral pharyngeal cartilage and the jaw joint. *Development* **130**, 1353-1365 (2003).

Miura,N., Wanaka,A., Tohyama,M., & Tanaka,K. MFH-1, a new member of the fork head domain family, is expressed in developing mesenchyme. *FEBS Lett.* **326**, 171-176 (1993).

Monaghan,A.P., Kaestner,K.H., Grau,E., & Schutz,G. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* **119**, 567-578 (1993).

Moore,A.W., McInnes,L., Kreidberg,J., Hastie,N.D., & Schedl,A. YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* **126**, 1845-1857 (1999).

Muhr,J., Andersson,E., Persson,M., Jessell,T.M., & Ericson,J. Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. *Cell* **104**, 861-873 (2001).

Murakami,R., Takashima,S., & Hamaguchi,T. Developmental genetics of the Drosophila gut: specification of primordia, subdivision and overt-differentiation. *Cell Mol. Biol. (Noisy-le-grand)* **45**, 661-676 (1999).

Murphy,K.C. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J. Bacteriol.* **180**, 2063-2071 (1998).

Muyrers,J.P., Zhang,Y., Testa,G., & Stewart,A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27**, 1555-1557 (1999).

Narita,T., Ishii,Y., Nohno,T., Noji,S., & Yasugi,S. Sonic hedgehog expression in developing chicken digestive organs is regulated by epithelial-mesenchymal interactions. *Dev. Growth Differ.* **40**, 67-74 (1998).

Narita,T., Saitoh,K., Kameda,T., Kuroiwa,A., Mizutani,M., Koike,C., Iba,H., & Yasugi,S. BMPs are necessary for stomach gland formation in the chicken embryo: a study using virally induced BMP-2 and Noggin expression. *Development* **127**, 981-988 (2000).

Naski,M.C., Colvin,J.S., Coffin,J.D., & Ornitz,D.M. Repression of hedgehog signalling and BMP4 expression in growth plate cartilage by fibroblast growth factor receptor 3. *Development* **125**, 4977-4988 (1998).

Neubuser,A., Koseki,H., & Balling,R. Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1. *Dev. Biol.* **170**, 701-716 (1995).

Newman,C.S., Grow,M.W., Cleaver,O., Chia,F., & Krieg,P. Xbap, a vertebrate gene related to bagpipe, is expressed in developing craniofacial structures and in anterior gut muscle. *Dev. Biol.* **181**, 223-233 (1997).

Newman,C.S. & Krieg,P.A. The Xenopus bagpipe-related homeobox gene zampogna is expressed in the pharyngeal endoderm and the visceral musculature of the midgut. *Dev. Genes Evol.* **209**, 132-134 (1999).

Newman,C.S. & Krieg,P.A. Xenopus bagpipe-related gene, koza, may play a role in regulation of cell proliferation. *Dev. Dyn.* **225**, 571-580 (2002).

Nicolas,S., Caubit,X., Massacrier,A., Cau,P., & Le Parco,Y. Two Nkx-3-related genes are expressed in the adult and regenerating central nervous system of the urodele Pleurodeles waltl. *Dev. Genet.* **24**, 319-328 (1999).

Nielsen,C., Murtaugh,L.C., Chyung,J.C., Lassar,A., & Roberts,D.J. Gizzard formation and the role of Bapx1. *Dev. Biol.* **231**, 164-174 (2001).

Nobrega,M.A., Ovcharenko,I., Afzal,V., & Rubin,E.M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).

Numata,K., Kanai,A., Saito,R., Kondo,S., Adachi,J., Wilming,L.G., Hume,D.A., Hayashizaki,Y., & Tomita,M. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.* **13**, 1301-1306 (2003).

Oh,S.P. & Li,E. The signalling pathway mediated by the type IIB activin receptor controls axial patterning and lateral asymmetry in the mouse. *Genes Dev.* **11**, 1812-1826 (1997).

Oh,S.P., Yeo,C.Y., Lee,Y., Schrewe,H., Whitman,M., & Li,E. Activin type IIA and IIB receptors mediate Gdf11 signaling in axial vertebral patterning. *Genes Dev.* **16**, 2749-2754 (2002).

Ohlsson,H., Karlsson,K., & Edlund,T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J.* **12**, 4251-4259 (1993).

Oliver,G., Wehr,R., Jenkins,N.A., Copeland,N.G., Cheyette,B.N., Hartenstein,V., Zipursky,S.L., & Gruss,P. Homeobox genes and connective tissue patterning. *Development* **121**, 693-705 (1995).

Osoegawa,K., Tateno,M., Woon,P.Y., Frengen,E., Mammoser,A.G., Catanese,J.J., Hayashizaki,Y., & de Jong,P.J. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**, 116-128 (2000).

Pabst,O., Zweigerdt,R., & Arnold,H.H. Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development* **126**, 2215-2225 (1999).

Patterson,K.D. & Krieg,P.A. Hox11-family genes XHox11 and XHox11L2 in xenopus: XHox11L2 expression is restricted to a subset of the primary sensory neurons. *Dev. Dyn.* **214**, 34-43 (1999).

Patterson,K.D., Drysdale,T.A., & Krieg,P.A. Embryonic origins of spleen asymmetry. *Development* **127**, 167-175 (2000).

Pennacchio,L.A. & Rubin,E.M. Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* **2**, 100-109 (2001).

Percival,A.C. & Slack,J.M. Analysis of pancreatic development using a cell lineage label. *Exp. Cell Res.* **247**, 123-132 (1999).

Peters,H., Wilm,B., Sakai,N., Imai,K., Maas,R., & Balling,R. Pax1 and Pax9 synergistically regulate vertebral column development. *Development* **126**, 5399-5408 (1999).

Pfaff,S.L., Mendelsohn,M., Stewart,C.L., Edlund,T., & Jessell,T.M. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309-320 (1996).

Pictet,R.L., Clark,W.R., Williams,R.H., & Rutter,W.J. An ultrastructural analysis of the developing embryonic pancreas. *Dev. Biol.* **29**, 436-467 (1972).

Ramalho-Santos,M., Melton,D.A., & McMahon,A.P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763-2772 (2000).

Reuter,R. & Scott,M.P. Expression and function of the homoeotic genes Antennapedia and Sex combs reduced in the embryonic midgut of Drosophila. *Development* **109**, 289-303 (1990).

Rindi,G., Ratineau,C., Ronco,A., Candusso,M.E., Tsai,M., & Leiter,A.B. Targeted ablation of secretin-producing cells in transgenic mice reveals a common differentiation pathway with multiple enteroendocrine cell lineages in the small intestine. *Development* **126**, 4149-4156 (1999).

Roberts,C.W., Shutter,J.R., & Korsmeyer,S.J. Hox11 controls the genesis of the spleen. *Nature* **368**, 747-749 (1994).

Roberts,D.J., Johnson,R.L., Burke,A.C., Nelson,C.E., Morgan,B.A., & Tabin,C. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* **121**, 3163-3174 (1995).

Roberts,D.J., Smith,D.M., Goff,D.J., & Tabin,C.J. Epithelial-mesenchymal signalling during the regionalization of the chick gut. *Development* **125**, 2791-2801 (1998).

Rodrigo,I., Hill,R.E., Balling,R., Munsterberg,A., & Imai,K. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* **130**, 473-482 (2003).

Rugh,R. The Mouse. (Oxford University Press, 1990).

Saito,H. The development of the spleen in the Australian lungfish, Neoceratodus forsteri Krefft, with special reference to its relationship to the "gastro"-enteric vasculature. *Am. J. Anat.* **169**, 337-360 (1984).

Saito,M., Iwawaki,T., Taya,C., Yonekawa,H., Noda,M., Inui,Y., Mekada,E., Kimata,Y., Tsuru,A., & Kohno,K. Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. *Nat. Biotechnol.* **19**, 746-750 (2001).

Sakiyama,J., Yokouchi,Y., & Kuroiwa,A. HoxA and HoxB cluster genes subdivide the digestive tract into morphological domains during chick development. *Mech. Dev.* **101**, 233-236 (2001).

Sandelin,A., Alkema,W., Engstrom,P., Wasserman,W.W., & Lenhard,B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91-D94 (2004).

Santini,S., Boore,J.L., & Meyer,A. Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res.* **13**, 1111-1122 (2003).

Schneider,A., Mijalski,T., Schlange,T., Dai,W., Overbeek,P., Arnold,H.H., & Brand,T. The homeobox gene NKX3.2 is a target of left-right signalling and is expressed on opposite sides in chick and mouse embryos. *Curr. Biol.* **9**, 911-914 (1999).

Schneider,A., Brand,T., Zweigerdt,R., & Arnold,H. Targeted disruption of the Nkx3.1 gene in mice results in morphogenetic defects of minor salivary glands: parallels to glandular duct morphogenesis in prostate. *Mech. Dev.* **95**, 163-174 (2000).

Schwartz,S., Zhang,Z., Frazer,K.A., Smit,A., Riemer,C., Bouck,J., Gibbs,R., Hardison,R., & Miller,W. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577-586 (2000).

Serfling,E., Lubbe,A., Dorsch-Hasler,K., & Schaffner,W. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO J.* **4**, 3851-3859 (1985).

Sharpe,J., Ahlgren,U., Perry,P., Hill,B., Ross,A., Hecksher-Sorensen,J., Baldock,R., & Davidson,D. Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science* **296**, 541-545 (2002).

Sharpe,J. Optical projection tomography as a new tool for studying embryo anatomy. *J. Anat.* **202**, 175-181 (2003).

Shi,Y. & Massague,J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).

Slack,J.M. Developmental biology of the pancreas. *Development* **121**, 1569-1580 (1995).

Smith,D.M. & Tabin,C.J. BMP signalling specifies the pyloric sphincter. *Nature* **402**, 748-749 (1999).

Smith,D.M., Nielsen,C., Tabin,C.J., & Roberts,D.J. Roles of BMP signalling and Nkx2.5 in patterning at the chick midgut-foregut boundary. *Development* **127**, 3671-3681 (2000a).

Smith,D.M., Grasty,R.C., Theodosiou,N.A., Tabin,C.J., & Nascone-Yoder,N.M. Evolutionary relationships between the amphibian, avian, and mammalian stomachs. *Evol. Dev.* **2**, 348-359 (2000b).

Spitz,F., Gonzalez,F., & Duboule,D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417 (2003).

Swaminathan,S., Ellis,H.M., Waters,L.S., Yu,D., Lee,E.C., Court,D.L., & Sharan,S.K. Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis.* **29**, 14-21 (2001).

Szymanski,M., Erdmann,V.A., & Barciszewski,J. Noncoding regulatory RNAs database. *Nucleic Acids Res.* **31**, 429-431 (2003).

Tanaka,M., Lyons,G.E., & Izumo,S. Expression of the Nkx3.1 homobox gene during pre and postnatal development. *Mech. Dev.* **85**, 179-182 (1999).

Theiler,K. The House Mouse: Development and Normal Stages from Fertilisation to 4 weeks of Age. (Springer-Verlag, Berlin, 1972).

Theiler,K. The House Mouse: Atlas of Embryonic Development. (Springer-Verlag, New York, 1989).

Theodosiou,N.A. & Tabin,C.J. Wnt signalling during development of the gastrointestinal tract. *Dev. Biol.* **259**, 258-271 (2003).

Thiel,G.A. & Downey,H. The development of the mammalian spleen, with special reference to its hematopoietic activity. *The American Journal of Anatomy* **28**, 279-339 (1921).

Thorvaldsen,J.L. & Bartolomei,M.S. Molecular biology. Mothers setting boundaries. *Science* **288**, 2145-2146 (2000).

Tissier-Seta,J.P., Mucchielli,M.L., Mark,M., Mattei,M.G., Goridis,C., & Brunet,J.F. Barx1, a new mouse homeodomain transcription factor expressed in cranio-facial ectomesenchyme and the stomach. *Mech. Dev.* **51**, 3-15 (1995).

Tremml,G. & Bienz,M. Homeotic gene expression in the visceral mesoderm of Drosophila embryos. *EMBO J.* **8**, 2677-2685 (1989).

Tribioli,C., Frasch,M., & Lufkin,T. Bapx1: an evolutionary conserved homologue of the Drosophila bagpipe homeobox gene is expressed in splanchnic mesoderm and the embryonic skeleton. *Mech. Dev.* **65**, 145-162 (1997).

Tribioli,C. & Lufkin,T. The murine Bapx1 homeobox gene plays a critical role in embryonic development of the axial skeleton and spleen. *Development* **126**, 5699-5711 (1999).

Wagner,K.D., Wagner,N., & Schedl,A. The complex life of WT1. *J. Cell Sci.* **116**, 1653-1658 (2003).

Walker,W.F. & Liem,K.F. Functional Anatomy of the Vertebrates: An Evolutionary Perspective. (Saunders College Publishing, New York, 1994).

Wasserman,W.W., Palumbo,M., Thompson,W., Fickett,J.W., & Lawrence,C.E. Human-mouse genome comparisons to locate regulatory sites. *Nat. Genet.* **26**, 225-228 (2000).

Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P., Antonarakis,S.E., Attwood,J., Baertsch,R., Bailey,J., Barlow,K., Beck,S., Berry,E., Birren,B., Bloom,T., Bork,P., Botcherby,M., Bray,N., Brent,M.R., Brown,D.G., Brown,S.D., Bult,C., Burton,J., Butler,J., Campbell,R.D., Carninci,P., Cawley,S., Chiaromonte,F., Chinwalla,A.T., Church,D.M., Clamp,M., Clee,C., Collins,F.S., Cook,L.L., Copley,R.R., Coulson,A., Couronne,O., Cuff,J., Curwen,V., Cutts,T., Daly,M., David,R., Davies,J., Delehaunty,K.D., Deri,J., Dermitzakis,E.T., Dewey,C., Dickens,N.J., Diekhans,M., Dodge,S., Dubchak,I., Dunn,D.M., Eddy,S.R., Elnitski,L., Emes,R.D., Eswara,P., Eyras,E., Felsenfeld,A., Fewell,G.A., Flicek,P., Foley,K., Frankel,W.N., Fulton,L.A., Fulton,R.S., Furey,T.S., Gage,D., Gibbs,R.A., Glusman,G., Gnerre,S., Goldman,N., Goodstadt,L., Grafham,D., Graves,T.A., Green,E.D., Gregory,S., Guigo,R., Guyer,M., Hardison,R.C., Haussler,D., Hayashizaki,Y., Hillier,L.W., Hinrichs,A., Hlavina,W., Holzer,T., Hsu,F., Hua,A., Hubbard,T., Hunt,A., Jackson,I., Jaffe,D.B., Johnson,L.S., Jones,M., Jones,T.A., Joy,A., Kamal,M., Karlsson,E.K., Karolchik,D., Kasprzyk,A., Kawai,J., Keibler,E., Kells,C., Kent,W.J., Kirby,A., Kolbe,D.L., Korf,I., Kucherlapati,R.S., Kulbokas,E.J., Kulp,D., Landers,T., Leger,J.P.,

Leonard,S., Letunic,I., Levine,R., Li,J., Li,M., Lloyd,C., Lucas,S., Ma,B., Maglott,D.R., Mardis,E.R., Matthews,L., Mauceli,E., Mayer,J.H., McCarthy,M., McCombie,W.R., McLaren,S., McLay,K., McPherson,J.D., Meldrim,J., Meredith,B., Mesirov,J.P., Miller,W., Miner,T.L., Mongin,E., Montgomery,K.T., Morgan,M., Mott,R., Mullikin,J.C., Muzny,D.M., Nash,W.E., Nelson,J.O., Nhan,M.N., Nicol,R., Ning,Z., Nusbaum,C., O'Connor,M.J., Okazaki,Y., Oliver,K., Overton-Larty,E., Pachter,L., Parra,G., Pepin,K.H., Peterson,J., Pevzner,P., Plumb,R., Pohl,C.S., Poliakov,A., Ponce,T.C., Ponting,C.P., Potter,S., Quail,M., Reymond,A., Roe,B.A., Roskin,K.M., Rubin,E.M., Rust,A.G., Santos,R., Sapojnikov,V., Schultz,B., Schultz,J., Schwartz,M.S., Schwartz,S., Scott,C., Seaman,S., Searle,S., Sharpe,T., Sheridan,A., Shownkeen,R., Sims,S., Singer,J.B., Slater,G., Smit,A., Smith,D.R., Spencer,B., Stabenau,A., Stange-Thomann,N., Sugnet,C., Suyama,M., Tesler,G., Thompson,J., Torrents,D., Trevaskis,E., Tromp,J., Ucla,C., Ureta-Vidal,A., Vinson,J.P., Von Niederhausern,A.C., Wade,C.M., Wall,M., Weber,R.J., Weiss,R.B., Wendl,M.C., West,A.P., Wetterstrand,K., Wheeler,R., Whelan,S., Wierzbowski,J., Willey,D., Williams,S., Wilson,R.K., Winter,E., Worley,K.C., Wyman,D., Yang,S., Yang,S.P., Zdobnov,E.M., Zody,M.C., & Lander,E.S. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).

Wells,J.M. & Melton,D.A. Vertebrate endoderm development. *Annu. Rev. Cell Dev. Biol.* **15**, 393-410 (1999).

Wessells,N.K. & Cohen,J.H. Early pancreas organogenesis: morphogenesis, tissue interactions and mass effects. *Dev. Biol.* **15**, 237-270 (1967).

Yamashita,H., ten Dijke,P., Huylebroeck,D., Sampath,T.K., Andries,M., Smith,J.C., Heldin,C.H., & Miyazono,K. Osteogenic protein-1 binds to activin type II receptors and induces certain activin-like effects. *J. Cell Biol.* **130**, 217-226 (1995).

Yassine,F., Fedecka-Bruner,B., & Dieterlen-Lievre,F. Ontogeny of the chick embryo spleen--a cytological study. *Cell Differ. Dev.* **27**, 29-45 (1989).

Yee,S.P. & Rigby,P.W. The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* **7**, 1277-1289 (1993).

Yokouchi,Y., Sakiyama,J., & Kuroiwa,A. Coordinated expression of Abd-B subfamily genes of the HoxA cluster in the developing digestive tract of chick embryo. *Dev. Biol.* **169**, 76-89 (1995).

Yoshiura,K.I. & Murray,J.C. Sequence and chromosomal assignment of human BAPX1, a bagpipe-related gene, to 4p16.1: a candidate gene for skeletal dysplasia. *Genomics* **45**, 425-428 (1997).

Yu,D., Ellis,H.M., Lee,E.C., Jenkins,N.A., Copeland,N.G., & Court,D.L. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A* **97**, 5978-5983 (2000).

Zhang,Y., Buchholz,F., Muyrers,J.P., & Stewart,A.F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123-128 (1998).

Appendix

Appendix 1: Recombineering

The targeting experiments involving the two *Bapx1*-containing PAC vectors were divided into two principal stages:

Targeting I: Removal of *sacB* from the pPAC4 vector backbone.

Targeting II: Modification of the *Bapx1* locus.

A1.1 Targeting I

Removal of the *sacB* gene from the pPAC4 vector backbone by homologous recombination had a dual purpose. Firstly, it would serve as a test of the system prior to attempting modification of the *Bapx1* locus. Secondly, it was required owing to the planned re-introduction of *sacB* into the PAC as part of a *cat* (chloramphenicol resistance gene)-*sacB* cassette at the *Bapx1* locus. The *cat-sacB* cassette would initially serve as a selectable marker to indicate correct targeting of the desired locus. It would subsequently allow for both selection and counter-selection when attempting replacement of the cassette with the cytotoxic gene (encoding the diptheria toxin receptor). The counterselectable marker *sacB* confers sucrose sensitivity to *E.coli*, killing cells in the presence of sucrose. Thus recombinants, having lost *sacB*, will be identified as sucrose resistant colonies that can be distinguished from spontaneous reversion mutants by testing for chloramphenicol sensitivity.

E.coli host DY380 cells containing the PAC vectors were transformed with a 118bp linear dsDNA fragment (50ng) which derived from the annealing of two 68 nucleotide overlapping primers which were subsequently 'filled-in' to yield dsDNA. The two primers each contained 50bp sequence homology immediately flanking *sacB* and were designed to replace the entire *sacB* coding region. In its place was incorporated the 18 nucleotide *I-Sce*I restriction site by including this 18bp sequence at the 3' end of the upstream primer and 5' end of the downstream primer (Figure A1.1). This modification placed a second *I-Sce*I site in the pPAC4 vector backbone immediately flanking the genomic insert, which could thus be separated from the

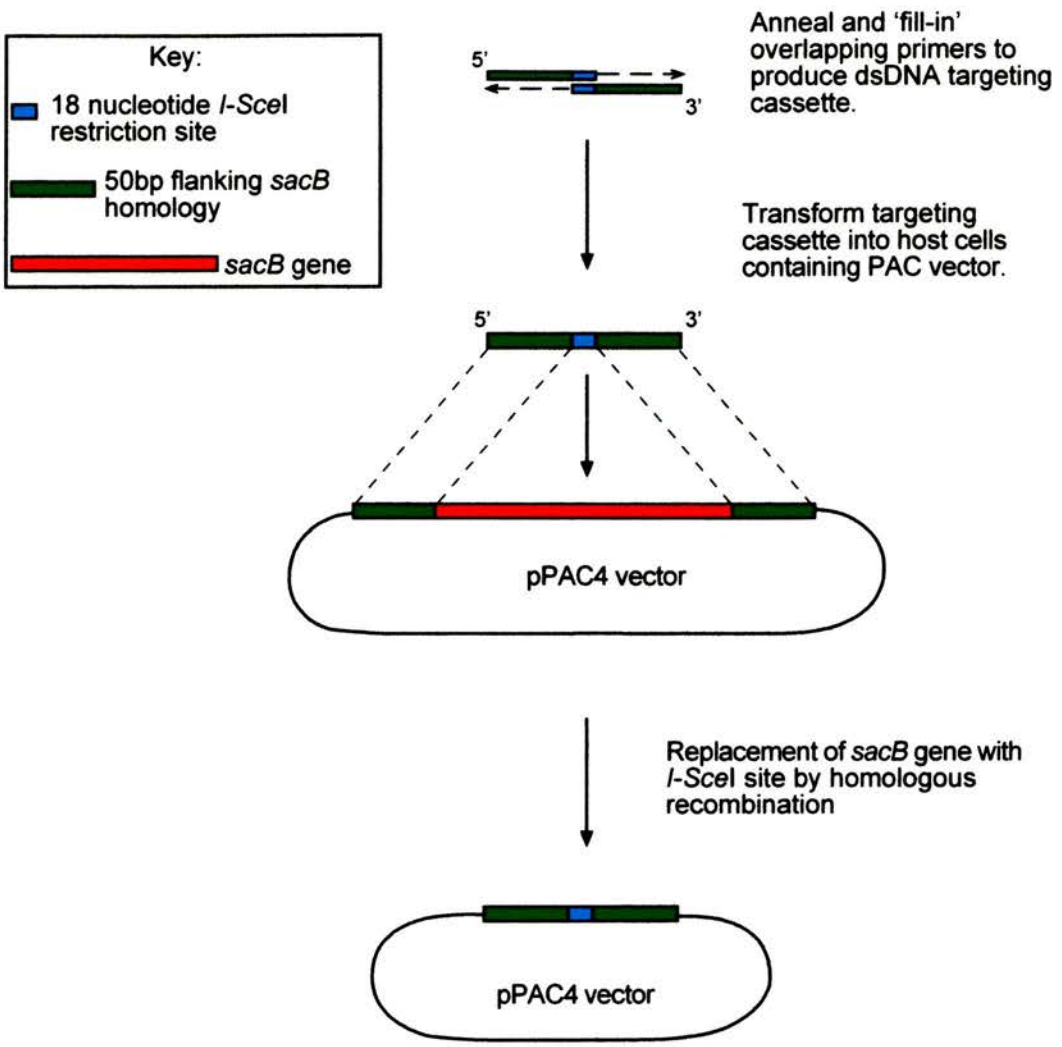


Figure A1.1: Key stages in the removal of the *sacB* gene from the pPAC4 vector backbone. Two overlapping oligonucleotides containing a common 18bp *I-Sce*I restriction site were annealed and DNA 'filled-in' to create a dsDNA targeting cassette. DNA sequence either side of the *I-Sce*I restriction site was homologous to pPAC4 vector sequence flanking the *sacB* gene. The targeting cassette was electroporated into host DY380 cells which provided functions to protect and recombine the linear cassette into the pPAC4 vector, replacing *sacB*.

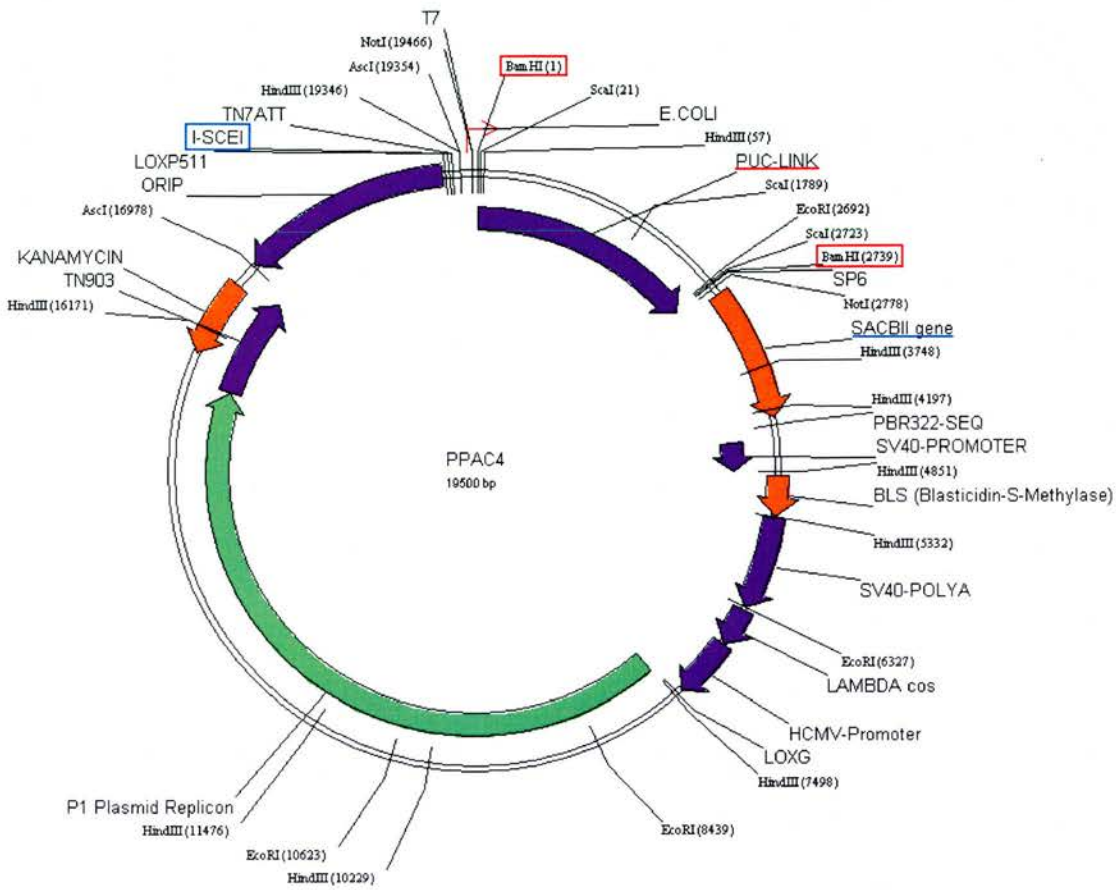


Figure A1.2: A map of the pPAC4 vector created by Eirik Frengen. During the original construction of the RPCI21 mouse genomic PAC library, the PUC-LINK stuffer fragment (underlined in red) was replaced by cloning partially *Mbo*I-digested female 129S6/SvEvTac mouse spleen genomic DNA between the *Bam*HI sites (boxed) (Osoegawa *et al.*, 2000). Diagram taken from the resources website of the laboratory in which the vector was created (http://bacpac.chori.org/). The *sacB* gene (underlined in blue) was replaced by insertion of an *I-Sce*I site additional to the one (boxed in blue) located on the other side of the genomic insert.

vector backbone by an *I-Sce*I digest. A map of the pPAC4 vector appears in Figure A1.2.

PCR analysis using a combination of primers to the targeted locus suggested that correctly targeted PAC clones were present and thus stage I, removal of *sacB* from the pPAC4 vector backbone, was successful (data not shown). It should be noted however that in some cases, pPAC4-derived *sacB* was detected with internal *sacB* primers following colony PCR yet no longer observed following purification of the PAC vector DNA away from the *E.coli* host genome. These results indicate that illegitimate recombination may have occurred such that the excised *sacB* gene has recombined into the *E.coli* DY380 host genome.

A1.2 Targeting II

Overnight cultures of DY380 cells containing the modified PAC vector (and negative for *sacB*) were induced and transformed with 100ng *cat-sacB* targeting cassette. The latter was generated by PCR using *cat-sacB* primers which included 51 additional nucleotides on each primer homologous to part of exon1 in the *Bapx1* gene. The PCR substrate was NC397 (*E.coli*) genomic DNA. Correct targeting to the *Bapx1* locus would result in the interruption of exon1 and replacement of 244 nucleotides with the *cat-sacB* fragment, as illustrated below in Figure A1.3.

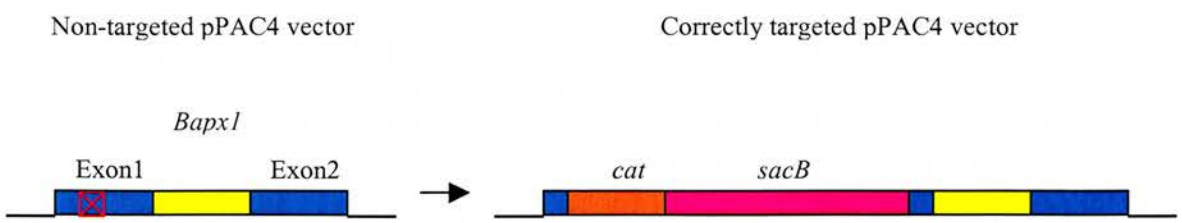


Figure A1.3: A cartoon depicting the targeting of a *cat-sacB* cassette to the *Bapx1* locus within the PAC vector. *Bapx1* exons are shaded blue and the intron is shaded yellow. The box containing a red cross identifies the 244bp *Bapx1* sequence replaced by insertion of the *cat-sacB* cassette.

A selection of potentially targeted colonies growing on chloramphenicol plates were subjected to PCR analysis using primers designed to identify the modified locus. Results indicated that the *cat-sacB* donor cassette had been integrated into either the PAC vector or the *E.coli* host (DY380) genome yet it was not correctly targeted to the *Bapx1* locus (data not shown). The same result was achieved following three separate attempts to modify the *Bapx1* locus. A selection of colonies identified as possessing the *cat-sacB* cassette was grown up overnight and PAC DNA harvested prior to digestion with *Asp718*I and Southern blot hybridisation with a *cat-sacB* probe. No signal was detected in any lanes containing digested PAC DNA (data not shown). These results are collectively accommodated by postulating incorporation of the *cat-sacB* fragment into the *E.coli* host genome, leaving the *Bapx1* locus in the PAC unmodified. The latter conclusion is further supported by experimental observations during stage I of the recombineering process.

A1.3 Conclusions and Discussion

It is not clear how either the pPAC4-derived *sacB* gene or NC397-derived *cat-sacB* with flanking *Bapx1* homology could have recombined into the genome of the DY380 cell. Neither is expected to possess significant homology to host sequences, yet the entire fragments appear to have relocated. Perhaps, in the case of the *cat-sacB* cassette, the selective pressure to acquire chloramphenicol resistance is great enough to drive illegitimate recombination. This explanation does not however hold for the excised *sacB* gene.

It is conceivable that the excision of the 1,421bp *sacB* gene is more readily permitted than the addition of 3,154bp into exon1 of *Bapx1*. Perhaps success is influenced by the percentage homology of the cassette in relation to the target locus. To this end, a significantly greater proportion (100/118bp = 84.7%) of the *sacB* targeting cassette is homologous to its target site than the *Bapx1* equivalent (102/3256bp = 3.1%). Along this line, it has recently been reported that targeting cassettes using 500bp flanking homology as opposed to the much shorter 30-50bp

yield considerably improved targeting efficiencies (Liu *et al.*, 2003). Nevertheless, this somewhat defeats the purpose of using recombineering technology since an advantage of recombineering is the ability to avoid cloning steps by using short stretches of homology for which oligonucleotide primers can readily be synthesised.

For both myself and others within the Human Genetics Unit, experiments were hampered by poor survival of cells and lower levels of recombination efficiency than expected based on the literature. Published recombination efficiencies have been in the region of 0.1% of surviving cells from a standard electroporation (Yu *et al.*, 2000). Although I calculated the transformation efficiency of host DY380 cells to be $2x10^7$ per μg electroporated DNA, generally fewer than 50 colonies survived on the selective plates and frequently less than 10 colonies were present. Illegitimate recombination is likely to be a rare event that would not normally be detected amongst the thousands of surviving colonies expected. In spite of these setbacks, it was the apparent inability to target the chosen locus that was the principal reason for adopting an alternative strategy to pursue our individual investigations.

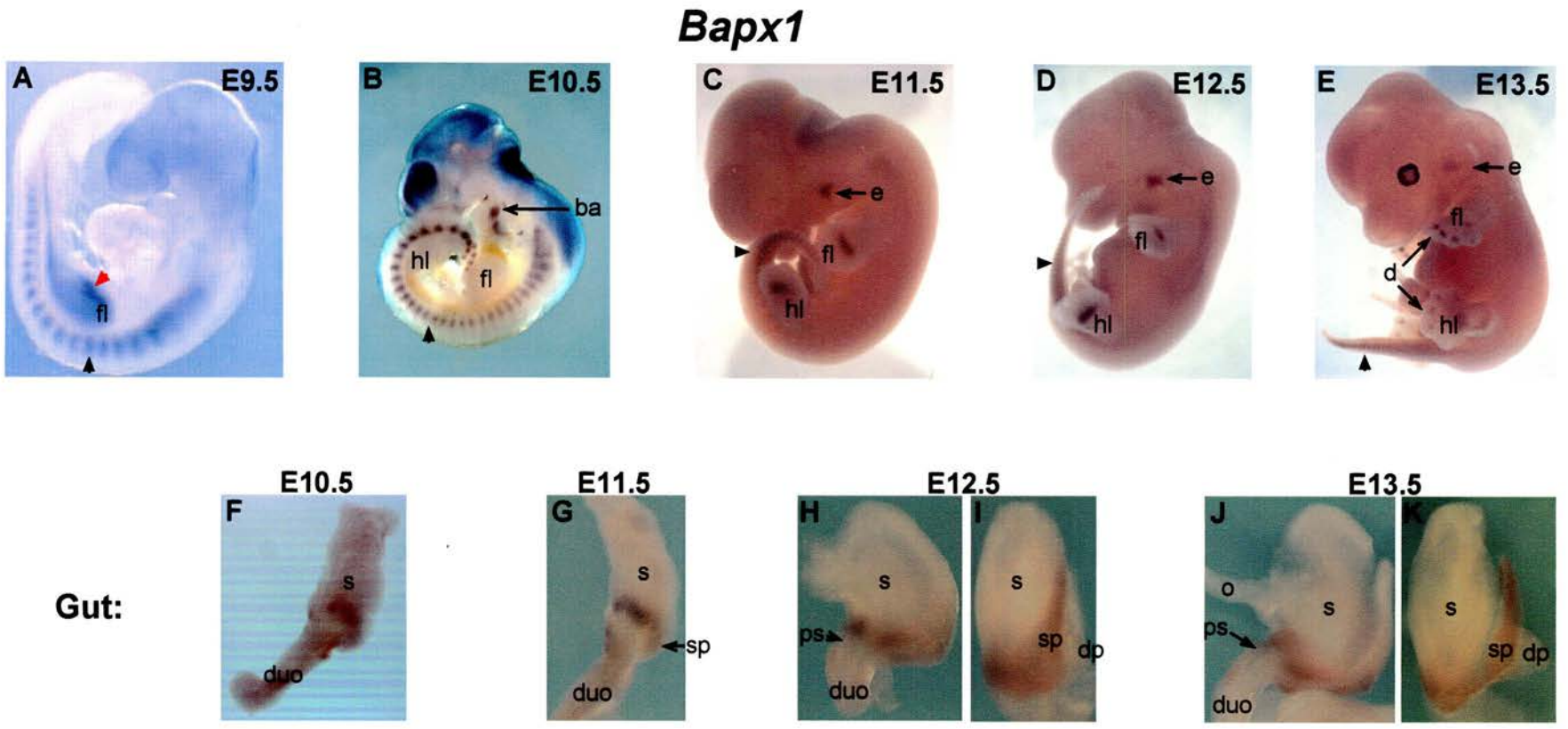


Figure A2: A panel to illustrate the wildtype *Bapx1* expression pattern as visualised following wholemount *in situ* hybridisation of whole embryos (**A-E**; left lateral view) and dissected guts (**F-H**, **J**; left lateral view; **I**, **K**; dorsal view) between E9.5 and E13.5 days of development. *Bapx1* is first detectable in the somites (black arrowheads) and is expressed in the lateral plate mesoderm surrounding the gut (red arrowhead) at E9.5. By E10.5 expression is detected externally in the first branchial arch (ba), later becoming confined to the middle ear, and by E11.5 *Bapx1* is detected in the cartilaginous condensations of the future fore- and hindlimb digits. Within the developing gut, *Bapx1* is expressed in the posterior stomach and the spleen. Images **B-E** and **F** were kindly provided by Dr. Robert Watson. Image **A** derives from Rodrigo *et al.*, 2003. To view *Bapx1* expression at E14.5 and E16.5 in sectioned embryos please refer to Tribioli *et al.*, 1997. ba, branchial arch; d, digits; dp, dorsal pancreas; duo, duodenum; e, ear; fl, forelimb; hl, hindlimb; o, oesophagus; ps, pyloric sphincter; s, stomach; sp, spleen.

Appendix 3

Line designation	Litter size and Number of embryos staining per litter									Average values	
	E8.5	E9.5	E10.5	E11.5	E12.5	E13.5	E14.5	E15.5	E16.5	No.	%
A266	–	–	13 0	14 4	8 2	11 5	13 3	13 1	–	12	26
A266.1	–	–	10 0	20 0	13 0	16 0	16 0	–	–	15	0
A266.2	9 0	–	11 0	16 0	10 1 12 0	7 0 6 2	14 1	12 0	–	11	17
A266.3	15 9	14 7	13 5	17 9	14 9	16 6	14 6	16 13	–	15	53
A266.4	–	–	14 0	15 8	15 9	5 1	15 5	15 0	–	13	42
A266.5	–	–	16 0	–	17 0	12 0	18 0	–	–	16	0
A267	–	12 5	14 9	14 7	11 6	12 7	10 5 8 4	–	–	12	53
~~A269~~											
A269.1	–	13 0	13 0	19 0	13 0	17 0	16 0	15 0	–	15	0
A269.2	–	21 1	19 6	14 2	11 4	15 5	14 3	13 2	2 1	14	26
A269.3	–	–	10 0	15 0	15 0	13 0	12 0	–	–	13	0
~~A269.4~~											
A269.5	–	14 6	15 8	1 0 14 8	17 6	15 9	20 11	17 6	–	14	48
~~A269.6~~											

Table A3: A table detailing the litter size of embryos harvested and number of embryos per litter staining positive for *LacZ* expression. Transgenic lines are listed to the left with DistB1 above the horizontal red division and ProxB below it. Average litter sizes and percentage of embryos staining for each line are listed to the right of the heavy black line. Developmental stages not tested are indicated by – and are not included in the analysis. Only stages at which reporter gene expression was detected were included in the calculation of percentage of embryos staining for a given line. Line A269 did not transmit the transgene and lines A269.4 and A269.6 could not be assayed due to failure to breed and infertility respectively.